DEVELOPMENT OF DEEP LEARNING METHODS FOR HEAD AND NECK CANCER DETECTION IN HYPERSPECTRAL IMAGING AND DIGITAL PATHOLOGY FOR SURGICAL GUIDANCE

A Dissertation
Presented to
The Academic Faculty

by

Martin Thomas Halicek

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Wallace H. Coulter Department of Biomedical Engineering

Emory University &
Georgia Institute of Technology
May 2020

DEVELOPMENT OF DEEP LEARNING METHODS FOR HEAD AND NECK CANCER DETECTION IN HYPERSPECTRAL IMAGING AND DIGITAL PATHOLOGY FOR SURGICAL GUIDANCE

Approved by:

Baowei Fei, PhD, Research Advisor
Department of Bioengineering
University of Texas at Dallas & University of Texas Southwestern Medical Center

James V. Little, MD
Department of Pathology & Laboratory Medicine
Emory University School of Medicine

Shella Keilholz, PhD, Academic Advisor
Department of Biomedical Engineering
Emory University & Georgia Institute of Technology

Richard Cameron, PhD
Department of Neuroscience & Regenerative Medicine
Medical College of Georgia, Augusta University

Erin Buckley, PhD
Department of Biomedical Engineering
Emory University & Georgia Institute of Technology

Date Approved: March 6, 2020

For my wife, our son, and our dogs

ACKNOWLEDGEMENTS

Firstly, I would like to acknowledge my advisor, Professor Baowei Fei, for the opportunity to perform the research in this thesis and his support during the four years I spent in his laboratory during my PhD studies. His enthusiasm for teaching and mentoring promotes a friendly and effective research environment that cultivates collaboration and achievement. He gave me crucial guidance using his vast expertise on medical imaging, allowed me independence to pursue my ideas, fully encouraged investigating emerging and cutting-edge methods, and helped me carefully plan my own path to ensure my success. I am grateful for not only the opportunity to work with him, but also for the opportunity to learn from him and all the members of our research group.

I would like to thank my thesis committee members: Dr. Shella Keilholz, Dr. Erin Buckley, Dr. James Little, and Dr. Richard Cameron. They offered guidance, time, and valuable feedback for improving the contents of this work, and I truly appreciate each of them for making the time and effort to help me during my PhD studies. I would also like to particularly thank Dr. Keilholz for serving as my co-advisor at Emory after our research group made the move to Texas. Her support and encouragement was greatly appreciated.

I would like to thank all the members of Dr. Fei's lab that I was truly fortunate to work so closely alongside, especially James Dormer and Dr. Maysam Shahedi for their friendship during my time. I was fortunate to have enjoyed collaborations and shared discussions that helped foster my enthusiasm for medical imaging, generate new ideas, and make the days even more enjoyable. I also want to acknowledge the collaborators from Las

Palmas, Spain that joined our research group, Dr. Himar Fabelo and Samuel Sarmiento, for their friendship and the exciting collaborations.

I would also like to thank the members of the surgical pathology department at Emory University Hospital Midtown: Andrew Balicki, Jacqueline Ernst, Tara Meade, Dana Uesry, and Mark Mainiero, for taking the time out of their work and their help collecting tissue specimens and answering my questions. Additionally, I would like to thank Dr. Xu Wang for helping coordinate the research at Emory University Hospital Midtown.

Lastly, none of the research work described in this dissertation would be at all possible without my family for their unconditional love and support: my wife, Maggie, my son, Maximilian, and our dogs, Scout, Scarlett, and Evie. My family was the most important component of my studies because they give my life meaning, which gives me motivation and inspiration to think and explore. My family and very close friends made this journey worthwhile. The opportunity to pursue a PhD in addition to my MD studies gave me the opportunity and flexibility to spend so much treasured time with my family and for that I am so fortunate and grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

2-NBDG	2-deoxy-2-[(7-nitro-2,1,3-benzoxadiazol-4-yl)amino]-D-glucose
2D	Two-dimensional
3D	Three-dimensional
ANN	Artificial Neural Network
AOTF	Acoustic-Optical Tunable Filter
AUC	Area Under the Curve Score
CCD	Charge-coupled Device
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DSC	Dice Similarity Coefficient
ES	Electromagnetic Spectrum
EUHM	Emory University Hospital Midtown
FAD	Flavin Adenine Dinucleotide
FCN	Fully-Convolutional Network
FOV	Field of View
FS	Frozen Section
FTC	Follicular Thyroid Carcinoma
Grad-CAM	Gradient Class-Activated Maps
H&E	Hematoxylin and Eosin
H&N	Head and Neck
Hb	Hemoglobin
HELICoiD	Hyperspectral Imaging Cancer Detection European project
HKM	Hierarchical K-Means
HPV	Human Papilloma Virus
HS	Hyperspectral
HSI	Hyperspectral Imaging
IPC	Intraoperative Pathologist Consultation
IRB	Institutional Review Board
kNN	k-Nearest Neighbor
LCTF	Liquid-Crystal Tunable Filter
LDA	Linear Discriminant Analysis
LED	Light-Emitting Diode
MC	Monte-Carlo
ML	Machine Learning
MLP	Multilayer Perceptron
MNG	Multinodular Goiter
MRI	Magnetic Resonance Imaging
MS	Multispectral
MSI	Multispectral Imaging
MTC	Medullary Thyroid Carcinoma
N	Normal Tissue Specimen

NAD	Nicotinamide Adenine Dinucleotide
NBI	Narrow Band Imaging
NIR	Near-Infrared
NPV	Negative Predictive Value
PCA	Principle Component Analysis
PLGA	Polymorphous Low-Grade Adenocarcinoma
PLS	Partial Least Squares
PPV	Positive Predictive Value
RBF	Radial Basis Function
RF	Random Forest
RGB	Red, green, blue
ROC	Receiver Operator Characteristic Curve
ROI	Region of Interest
SCC	Squamous Cell Carcinoma
sO2	Oxygen Saturation
SVM	Support Vector Machines
SWIR	Short-wavelength Infrared
T	Tumor Tissue Specimen
TN	Tumor-Normal Margin Tissue Specimen
TRE	Target Registration Error
US	Ultrasound
VNIR	Very Near-Infrared
WSI	Whole-Slide Images

SUMMARY

Surgeons performing routine cancer resections utilize palpation and visual inspection, along with time-consuming microscopic tissue analysis, to ensure removal of cancer. Despite this, inadequate surgical cancer margins are reported for up to 10-20% of head and neck squamous cell carcinoma (SCC) operations. There exists a need for surgical guidance with optical imaging to ensure complete cancer resection in the operating room. The objective of this dissertation is to evaluate hyperspectral imaging (HSI) as a non-contact, label-free optical imaging modality to provide intraoperative diagnostic information. For comparison of different optical methods, autofluorescence, RGB composite images synthesized from HSI, and two fluorescent dyes are also acquired and investigated for head and neck cancer detection. A novel and comprehensive dataset was obtained of 585 excised tissue specimens from 204 patients undergoing routine head and neck cancer surgeries.

The first specific aim of this thesis used SCC tissue specimens to determine if HSI has potential to perform surgical guidance in the challenging task of head and neck SCC detection. Currently, during SCC resections, several intraoperative pathology consultations are performed to investigate the excised tissues microscopically, which can take 20 to 45 minutes each. Therefore, it is hypothesized that, if proven reliable, HSI could reduce time and provide quantitative cancer predictions. Deep learning algorithms have demonstrated success at image classification and segmentation tasks by extracting relevant features from extremely large training datasets to allow generalizable learning and classification ability. In the first specific aim of this thesis, state-of-the-art deep learning algorithms were developed for SCC detection at the registered cancer margin in excised tissue specimens.

The ground truth for the gross-level images was obtained from histology, which was demonstrated to be effective up to 2 mm, and a method to estimate cancer detection performance at several distances from the cancer margin was developed. Investigating all 293 tissue specimens from 102 patients with SCC, HSI detected conventional type SCC with a median AUC score of 85% at 2 mm from the cancer margin. Evaluating performance of HSI by anatomical location demonstrated good SCC detection results in regions of the upper aerodigestive tract, such as the larynx (85% AUC score), oropharynx (95% AUC score), hypopharynx (78% AUC score), and nasal cavity (93% AUC score) at distances up to 2 mm from the cancer margin. To understand the ability of HSI for SCC detection, the most salient spectral features were calculated using gradients and correlated with known cancer physiology signals, notably oxygenated and deoxygenated hemoglobin. In conventional type SCC detection with HSI, the organ that produced the lowest results was the tongue, which may have been confounded by myoglobin signals. In HPV+ type SCC, autofluorescence imaging detected in tonsillar tissue with 91% AUC score, outperforming other modalities. The results of this aim indicate HSI may hold potential for guiding SCC resections.

The second specific aim of this investigation was to evaluate HSI for tumor detection in thyroid and salivary glands. For comparing HSI to traditional RGB imagery, RGB images were synthesized from HSI using the spectral response curves of the human eye. Using deep learning and our dataset of 200 thyroid tissue specimens from 76 patients with thyroid cancer, HSI detected thyroid tumors with 86% average AUC score, which outperformed fluorescent dyes and autofluorescence. However, HSI-synthesized RGB imagery significantly outperformed HSI with 90% AUC score. To understand why this was

possible for thyroid tumors, the most salient spectral features were calculated using gradients and indicated that for thyroid tumors the visible spectrum is sufficient. The results overall showed that HSI along with other modalities can be used for developing thyroid tumor predictive algorithms successfully with deep learning. In salivary gland tumors, HSI demonstrated the best performance for tissue specimens from 3 patients with parotid gland tumors (92% AUC score), and autofluorescence performed best for tissue specimens from 3 patients with tumors of other salivary glands (80% AUC score).

In the process of surgical pathologist consultations, pathologists interpret many histological slides, and we hypothesize that a computer aided diagnostic tool can be created for assistance. The final aim of this investigation was to develop deep learning algorithms for head and neck cancer detection in hundreds of digitized histology slides made from the excised tissue specimens in this study. Slides containing any amount of SCC or thyroid carcinoma can be distinguished from entirely normal slides with 94% and 99% AUC scores, respectively. Additionally, the results demonstrated that SCC and thyroid carcinoma can be detected and localized within digital whole-slide histology images with 92% and 95% AUC scores, respectively. The results of this aim indicate deep learning and digital histology could aid surgical pathologists.

The outcomes of this thesis work demonstrate that HSI and deep learning methods could aid surgeons and pathologists in detecting head and neck cancers. Future steps necessary for streamlining deployment for potential clinical translation of HSI are outlined, including moving from ex-vivo to in-vivo acquisitions. Additionally, future steps needed for translation of deep learning for assisting surgical pathologists are outlined. In conclusion, the methods developed and presented for HSI and histological imaging are

valuable for the broader, medical optical imaging field and serve as an example for future studies and emerging modalities for optical surgical guidance.

CHAPTER 1. INTRODUCTION

1.1 Background and Motivation

Estimated cancer incidence world-wide in 2012 for cancers of the oral/nasal cavity, nasopharynx, oropharynx, hypopharynx, larynx, and thyroid combined was 1.4 million people newly diagnosed. This corresponds with an age-standardized incidence rate of about 25 men out of 100,000 and 14 women out of 100,000 diagnosed in 2012 alone with these types of cancer [1]. Approximately 90% of cancer at sites including the lips, gums, mouth, palate, and anterior two-thirds of the tongue are squamous cell carcinoma (SCC) [2]. Significant risk factors for the development of these types of cancer are tobacco consumption, alcohol consumption, and infection with human papillomavirus (HPV) [3]. There were over 800,000 cancer deaths in 2012 from these types of cancer [1]. The diagnostic procedure of SCC typically involves physical examination and surgical evaluation by a physician, tissue biopsy, and diagnostic imaging, such as PET, MRI, or CT. The primary treatment for many cases of resectable SCC involves surgical cancer resection with potential adjuvant therapy, such as chemotherapy or radiation. For cases where routine surgery is the primary treatment modality, successful surgical cancer resection is required to prevent local disease recurrence and promote disease-free survival [4].

The most common practice during surgery to identify cancer margins is frozen section analysis, which can take 20 to 45 minutes [5-7]. When determining negative and positive cancer margins, most studies on oral cavity and aerodigestive tract cancers typically use a threshold of normal tissue beyond the tumor site greater than 5 mm to define

margin adequacy [6]. Studies comparing disease recurrence using the 5 mm margin definition found that inadequate margins carry increased locoregional recurrence rates (30-55%), compared to adequate, i.e. negative, margin recurrence rates (12-18%) [8-10]. Moreover, disease recurrence greatly affects likelihood for additional surgeries, reduced quality of life, and increased mortality rates [11]. Five recent studies, comprising 1,200 patients, found that positive cancer margins significantly reduce disease-free patient survival rates, values reported from 7-52%, compared with disease-free survival rates of 39-73% for cases where negative margins are achieved [8-10, 12, 13].

Hyperspectral imaging (HSI) is a non-contact and label-free optical imaging modality, which acquires a series of 2D images at hundreds of discrete wavelengths to capture both spatial information and a spectral signature of the target. Preliminary research demonstrates that HSI has potential for providing diagnostic information for a myriad of diseases, including anemia, hypoxia, cancer detection, skin lesion and ulcer identification, urinary stone analysis, enhanced endoscopy, and many potential others in development [14-25]. Supervised machine learning and artificial intelligence algorithms have demonstrated the ability to classify images after being allowed to learn features from training or example images. One such deep learning method, convolutional neural networks (CNNs), has demonstrated astounding performance at image classification tasks due to the capacity for robust handling of training sample variance and ability to extract meaningful features from large training data sizes [26, 27].

The need for an imaging modality that can perform diagnostic prediction could potentially aid surgeons with near real-time guidance during intra-operative cancer resection. If proven to be reliable and generalizable, this method could help provide intra-

operative diagnostic information beyond palpation and visual inspection to the surgeon's resources, perhaps enabling surgeons to achieve more accurate resections and biopsies, or as a computer-aided diagnostic tool for physicians diagnosing and treating these types of cancer.

1.1.1 Preliminary Studies

Hyperspectral imaging has been applied successfully to a number of biomedical challenges. Preliminary research demonstrates that HSI has potential for providing diagnostic information for diseases, such as anemia and hypoxia, for example determining tissue oxygenation and altitude induced hypoxia [28]. Additionally, HSI has demonstrated potential for in-vivo and ex-vivo cancer detection, such as identifying residual cancer in the tumor bed from a multi-institute intraoperative brain cancer study and colon cancer identification [15, 18, 19, 22]. Moreover, HSI has shown advancements in identification of skin lesions, diabetic ulcers, and urinary stone subtyping [25, 29]. Because of the demonstrated ability of HSI for identifying diseases, we chose to implement it in our study for cancer detection in surgical specimens from head and neck cancer, a unique application of HSI that has not been explored before, and to construct an extensive patient database to address these problems. Before applying the methods to human surgical specimens, the work was first demonstrated by our lab in several animal models.

A previous Ph.D. student (G. Lu) also advised by Professor Fei developed the imaging protocol for SCC detection in a mouse model using HSI. Lu and Fei were the first to publish a comprehensive review on the use of HSI for biomedical application, which has been cited nearly 900 times [14], indicating the surge of research in the field. Specifically,

a subcutaneous cancer model was made using a xenograft of HNSCC cell line into mice, and a support vector machines (SVM) classifier was implemented with 89% accuracy (n=12 mice) when HSI are acquired through the skin on the back of the mouse [30]. Afterwards, intraoperative HSI were acquired after skin resection and tumor exposure in the same mouse model, and using an active-contour classifier achieved results of 97% accuracy for tumor detection, cross-validating with only 5 mice [31]. Next, a mouse model of neoplastic 4NQO-induced tongue carcinogenesis was adopted, and SVM classification of neoplastic versus normal tissue yielded an AUC of 0.84 for in-vivo HSI and 0.86 AUC for *ex-vivo* HSI, again using cross-validations [32]. Lastly, Lu et al. began the *ex-vivo* HSI head and neck cancer surgical specimen study in collaboration with Emory University Hospital Midtown. Prior to this thesis, Lu et al. published a single preliminary work on human, ex-vivo cancer detection using HSI classified with SVM in limited regions of interest (ROIs) from 6 thyroid and 20 HNSCC non-registered patients' samples that achieved 86% accuracy for inter-patient validation and 91% accuracy for intra-patient validation, using cross-validations [33]. These works were done to establish the potential of HSI in a mouse model and to begin the process of evaluating HSI for *ex-vivo* human specimens. However, much more work remains to establish ground-truth registration for whole-tissue mapping and validation, complete investigation of more sophisticated and robust classification methods on a significantly larger dataset of patients, and to compare classification potential of HSI to other optical modalities and digitized histological images.

1.1.2 Significance and Innovation

As discussed, in routine cases of resectable head and neck, complete removal of the cancer is required to achieve successful disease-free survival and reduce local disease

recurrence [8-10, 12, 13]. Hyperspectral imaging, a non-contact, non-ionizing, and label-free optical imaging method, has demonstrated diagnostic potential that may aid surgeons with near real-time guidance during intra-operative cancer resection. If proven to be reliable and generalizable, we hypothesize the proposed method would add quantitative imaging to extend surgeons' resources beyond normal palpation and visual inspection, becoming a candidate for clinical translation to the operating room.

The variability of HSI data is substantial both within and between patients and, therefore, must be calibrated to compare patients. Though classes of tissue may show significant differences on average, the large standard deviations of spectral signatures from classes of tissue require machine learning to make predictions about tissue class. Despite calibration, the machine learning algorithms implemented must be sufficiently robust to handle the variation of spectral signatures from different tissues of different anatomical sites that may occur within one type of carcinoma. To overcome these challenges, we propose utilizing deep learning with deep CNNs and hypothesize the results will solve these challenges to a clinically acceptable degree, performing well in a task where traditional, regression-based methods are limited. For this task, image registration is essential for creating an accurate ground truth to allow validation of deep learning for cancer detection in ex-vivo tissues, with potential for HSI to serve as intraoperative surgical guidance. Additionally, inspecting correct and incorrect classification results, we propose to analyze the weights and activations from the deep learning methods to reveal which parts of the optical spectra was used for making the decision; this will allow investigation inside the "black-box" of deep learning and correlation with cancer pathophysiology. Altogether, these results and developments will facilitate thorough investigation of HSI-based cancer

detection, and serve as a standard for ex-vivo gross-level imaging studies, within all areas of optical modalities.

Prior to this thesis study, for head and neck cancers, HSI has been demonstrated to have benefit in the animal model (n=12 mice), and was investigated in human surgical specimens with only a limited sample size (n=26 patients). This work aims to address the challenges and difficulties of cancer detection using human surgical specimens from head and neck cancers using a substantially large patient database (n=204), a thorough investigation of machine learning methods, and comparison of HSI to other optical imaging modalities. Evaluating HSI-based cancer detection on an extensive and comprehensive *ex-vivo* patient database is an essential step to determine if the approach can be viable for clinical translation to intraoperative surgical guidance.

During surgical cancer resection, fresh frozen specimens are sent to the surgical pathology department to provide feedback on cancer margin status during surgery. Identifying cancer microscopically during intraoperative consultations can be challenging because of histological preparation and processing, time constraints, and the sheer volume of slides to make a final diagnosis. While we hypothesize that HSI could one day play a role supplying quantitative analysis to surgeons prior to resection, there is also a need for cancer detection in histopathological specimens. We hypothesize that a computer-aided diagnostic tool could be developed for assisting pathologists to quantitatively identify cancer is digitized histology slides made from surgical specimens.

To the best of our knowledge, this is the first work to identify head and neck cancer on histological images, which is only possible because of our extensive patient dataset.

Moreover, the results and developments of this work will establish a benchmark in the field for future work on computer aided diagnostics using histology. The major results and conclusions drawn from this thesis will help define the role and potential of artificial intelligence in the field of surgical guidance and digital pathology.

1.2 Specific Aims

The objective of this work is to evaluate hyperspectral imaging (HSI) as a non-contact, optical technique to further provide diagnostic information, for potential translation into the operating room. We hypothesize that label-free HSI will outperform autofluorescence, RGB composites, and fluorescent dye-based optical methods for the detection of head and neck cancer in ex-vivo fresh surgical tissue of human patients. In this thesis, we detail the HSI dataset obtained from 204 patients, a novel and comprehensive HSI database, and perform analysis using state-of-the-art deep learning methods, allowing thorough and complete investigation of cancer detection in ex-vivo surgical specimens of head and neck cancer. The results and outcomes of this comprehensive work demonstrate potential viability for clinical translation into intra-operative guidance systems.

<u>Specific Aim 1</u>: Acquire a novel and extensive database of HSI from ex-vivo specimens obtained from H&N cancer patients undergoing surgery and develop and evaluate advanced image processing techniques.

(**a)** Construct a novel, comprehensive ex-vivo HSI database of 204 patients with head and neck cancer, including SCC and thyroid carcinomas. The dataset is investigated and detailed: patient demographics, clinical descriptors, and surgical pathology information. Imaging protocols to acquire reflectance broadband HSI, autofluorescence,

two dye-based fluorescence techniques, and standard RGB imaging. (**b)** A method for preparing accurate and reliable histopathological gold standards for HSI using a pipeline of rigid and deformable image registration is developed and evaluated. Serial sectioning of a few specimens is performed to understand variation of the cancer margin with depth. **(c)** Develop and analyze processing techniques with comparison of machine learning methods for the HSI data, including assessing image normalization techniques and performance degradation due to specular glare.

Impact: This aim establishes the data necessary for this investigation of HSI, which will draw conclusions about clinical potential. The processing and registration methods developed in this aim allow for inter-patient machine learning experiments and quantify the systematic error in using histology as the ground truth for gross-level ex-vivo tissue specimens. Investigation of specular glare and the variation of the cancer margin provide more information to facilitate thorough investigation of HSI-based cancer detection, and serve as a standard for future ex-vivo gross-level imaging studies with optical modalities.

Specific Aim 2: Develop deep learning algorithms to classify head and neck cancer margins using HSI and optical imaging.

(a) Develop convolutional neural network algorithms for classification of optical spectral properties of ex-vivo tissue specimens, using HSI, autofluorescence, RGB composite imaging, 2NBDG dye, and proflavin dye-based imaging. **(b)** Perform cancer detection in both whole tumor-only and normal-only tissues for baseline, ideal-case performance. Perform cancer detection at the registered tumor-normal margin and erode distances from the margin to identify how distance affects performance. The described

experiments will be performed in fully-independent patient groups for all 102 SCC patients and 82 thyroid cancer patients in the HSI database. Compare performance of optical imaging modalities. Additionally, intra-patient experiments will be performed, which simulates preliminary augmented surgical approach. **(c)** Investigate correct classification results to analyze CNN weights and activations from deep learning and correlate with cancer pathophysiology.

Impact: Expected results will demonstrate potential of HSI for intraoperative cancer detection and surgical guidance. Provide support for the hypothesis that HSI has a unique ability compared to optical imaging methods to capture full diagnostic, optical spectra that comprise biological properties of malignant neoplasms in multiple cancer types at difference anatomical locations. The comparison of optical imaging modalities for disease classification will guide which techniques have potential to be translated into intra-operative guidance. Correlate CNN activation weights with theoretical optical features allows seeing inside the "black-box" of deep learning and provide physiological rationale.

Specific Aim 3: Compare HSI outcomes to histology-based methods. Develop and evaluate deep learning classifiers for head and neck cancer detection on digitized histological images.

(a) Determine if HSI can provide sub-component classification for normal and cancerous tissues, similar to histological analysis, which provides more detail than just binary classification. **(b)** Using the digitized histological dataset of surgical specimens, develop CNN-based methods for classification. Implement and evaluate histological cancer detection for SCC and thyroid carcinoma in digitized whole-slide images. Develop

color and stain invariant approach for robustness. The objective is to create methods for (1) localizing the cancer within each slide and delineating with heat-maps, and (2) identifying slides with cancer from entirely normal slides. **(c)** Perform gradient tracing analysis to understand salient features from CNN-based decision making with clinical rationale. **(d)** Establish limitations of the proposed method.

Impact: This will be the first work to identify head and neck SCC on histological images using deep learning. Not only will this serve as a basis in the field for future work on SCC detection using histology, but it also provides a benchmark for comparison of the ability of HSI-based cancer detection methods, which use the exact same specimens.

1.3 Thesis Overview

In this thesis work, the use of hyperspectral imaging was investigated for detection of excised human head and neck cancer specimens. Chapter 2 reviews the background and current literature available for hyperspectral imaging. Chapters 3-7 are reproductions of peer-reviewed, first-author journal manuscripts. Chapter 8 concludes the thesis with a summary of key findings, impact, and future research directions.

Chapter 2 is an extensive literature review on the studies utilizing HSI for cancer detection. It is a reproduction of the peer-reviewed manuscript, entitled "*In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: Revealing the invisible features of cancer*", by Martin Halicek et al. 2019, which was published in the journal *Cancers*.

Chapter 3 is a reproduction of the peer-reviewed manuscript, entitled "*Hyperspectral imaging for head and neck cancer detection: Specular glare and variance of the tumor margin in surgical specimens*", by Martin Halicek et al. 2019, which was published in the *Journal of Medical Imaging* and describes machine learning, processing, and ground truth methods. It also contains excerpts from a conference proceedings by Martin Halicek et al. 2018 on image registration.

Chapter 4 is a reproduction of the peer-reviewed manuscript, entitled "*Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning*", by Martin Halicek et al. 2019, which was published in the journal *Cancers*. This work details the complete investigation of HSI for SCC analysis, using all the SCC patients collected for this thesis divided into two cohorts, and compares the results of HSI with other optical imaging modalities.

Chapter 5 is a reproduction of the peer-reviewed manuscript, entitled "*Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning*", by Martin Halicek et al. 2020, which was published in the journal *Biomedical Optics Express*. This work details the complete investigation of HSI for thyroid and salivary gland tumor analysis, using all the thyroid and salivary gland tumor patients collected for this thesis, and compares the results of HSI with other optical imaging modalities.

Chapter 6 is a reproduction of the peer-reviewed manuscript, entitled "*Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural*

networks", by Martin Halicek et al., which was published in the *Journal of Biomedical Optics*. It describes the use of HSI for sub-classification of normal tissue components and sub-classification of different thyroid tumors.

Chapter 7 is a reproduction of the peer-reviewed manuscript, entitled "*Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks*", by Martin Halicek et al. 2019, which was published in the Nature journal *Scientific Reports*. It describes the application of cancer detection in the digitized whole-slide histological images prepared from the same tissue specimens used in the above HSI studies and previous chapters.

Chapter 8 is a summary of the key findings, limitations, future directions, and impacts of the research involved in this thesis work.

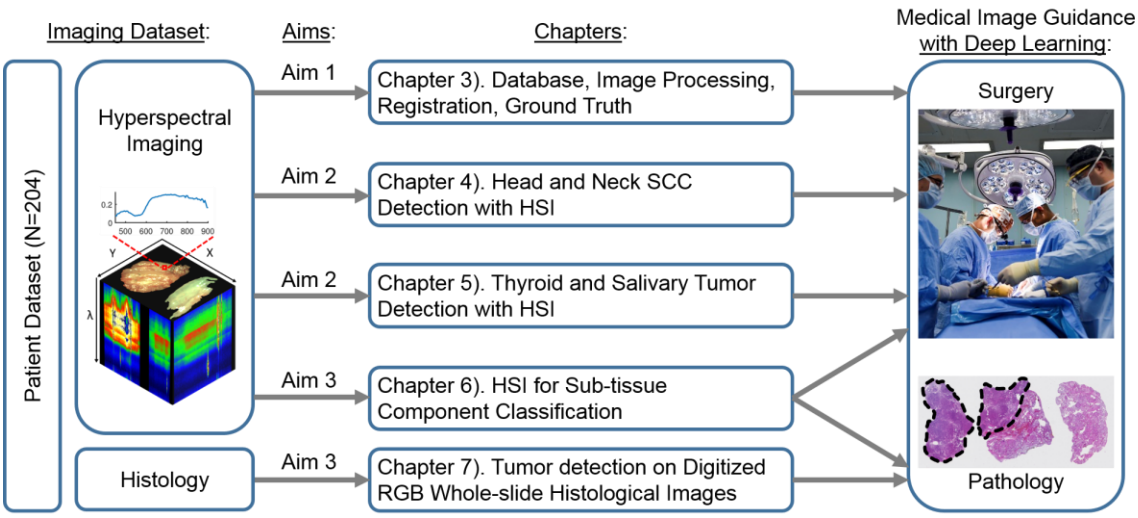

Figure 1-1. Flowchart of the dissertation.

CHAPTER 2. LITERATURE REVIEW: HYPERSPECTRAL IMAGING FOR CANCER ANALAYSIS

In this first chapter, a current literature review of the state of hyperspectral imaging for cancer detection is provided. In contrast to conventional optical imaging modalities, hyperspectral imaging (HSI) can capture much more information from a certain scene, both within and beyond the visual spectral range (from 400 to 700 nm). This imaging modality is based on the principle that each material provides different responses to light reflection, absorption, and scattering across the electromagnetic spectrum. Due to these properties, it is possible to differentiate and identify the different materials/substances presented in a certain scene by their spectral signature. Over the last two decades, HSI has demonstrated potential to become a powerful tool to study and identify several diseases in the medical field, being a non-contact, non-ionizing, and a label-free imaging modality. In this literature review, the use of HSI as an imaging tool for the analysis and detection of cancer is presented. The basic concepts related to this technology are detailed. The most relevant, state-of-the-art studies that can be found in the literature using HSI for cancer analysis are presented and summarized, both in-vivo and ex-vivo. Lastly, we discuss the current limitations of this technology in the field of cancer detection, together with some insights into possible future steps in the improvement of this technology.

2.1 Introduction

Hyperspectral imaging (HSI), also referred to as imaging spectroscopy, integrates conventional imaging and spectroscopy methods to obtain both spatial and spectral information of a scene [34]. Unlike conventional RGB (red, green, and blue) imaging, which only captures three diffuse Gaussian spectral bands in the visible spectrum (e.g., 380–740 nm), HSI increases the amount of data acquired beyond the capabilities of the

human eye [35]. Hyperspectral (HS) sensors measure the aggregate signal of reflected, absorbed and emitted radiance at specific wavelengths of the material that is being observed. These sensors are capable of capturing a very large number of contiguous spectral bands (also called spectral wavelengths or spectral channels) across the electromagnetic spectrum (ES). Each pixel of an HS image, or average over a region of interest, has associated a specific vector of radiance values, commonly called spectral signature [34]. Image processing algorithms make use of these spectral signatures to automatically differentiate the materials observed by the sensor at each pixel [36]. These methods rely on the basis that different molecular compositions of each material present in nature have different responses to the incident light [34].

HSI is a promising non-invasive and non-ionizing technique that supports rapid acquisition and analysis of diagnostic information in several fields, such as remote sensing [37, 38], archeology [39, 40], drug identification [41, 42], forensics [43-45], defense and security [46, 47], agriculture [48, 49], food safety inspection and control [50-52], among many others. Particularly, several studies can be found in the literature where HSI is applied to different medical applications [14, 53, 54]. It has been proven that the interaction between the electromagnetic radiation and matter carries useful information for medical diagnostic purposes [14]. As an alternative to other existing technologies for diagnosis, one of the strengths offered by HSI is being a completely non-invasive, non-contact, non-ionizing and label-free sensing technique. In medical applications, this technology has been employed in several different areas such as blood vessel visualization enhancement [55, 56], intestinal ischemia identification [57], oximetry of the retina [58-60], estimation of the cholesterol levels [61], chronic cholecystitis detection [62], histopathological tissue analysis, diabetic foot [63], etc. In recent years, medical HSI has started to achieve promising results with respect to cancer detection through the utilization of cutting-edge machine learning algorithms and increased modern computational power [14, 53].

In this literature review, the basis of the different HS acquisition systems is explained, as well as their applications to the different tissue sample analysis, followed by a brief presentation of the most common data processing approaches employed to process HS information. In addition, a brief introduction to the optical properties of cancer tissues and the current investigations performed in the literature to study and analyze in-vivo and ex-vivo cancer tissue samples using HSI are presented.

2.2 Hyperspectral Imaging

2.2.1 Hyperspectral Image Acquisition Systems

HSI sensors generate a three-dimensional (3D) data structure, called HS cube, where the spatial information is contained in the first two dimensions, while the third dimension encompasses the spectral information. Figure 2-1 shows the information structure of an HS cube. On one hand, each pixel of the HS image contains a full spectral signature of length equal to the number of spectral bands of the HS cube. On the other hand, a gray scale image of the captured scene can be obtained using any of the spectral bands that displays the spatial information provided by the image sensor at a particular wavelength.

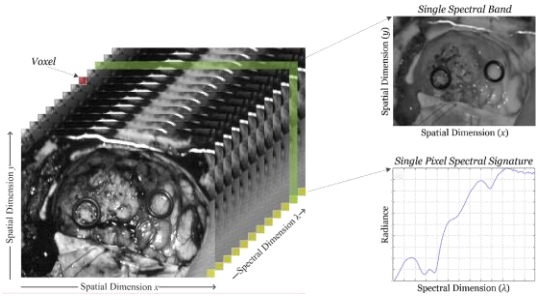

Figure 2-1. Hyperspectral imaging data. Basic structure of a hyperspectral imaging (HSI) cube, single band representation at a certain wavelength and spectral signature of a single pixel.

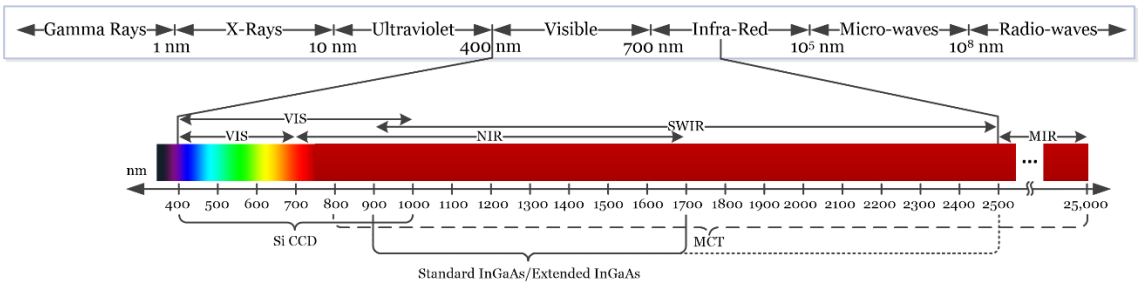

Figure 2-2. Electromagnetic spectrum. HSI is commonly employed between the visible and the medium-infrared range.

Depending on the type of sensor employed, HSI cameras will cover different spectral ranges. Figure 2-2 shows the partitioning of the entire electromagnetic spectrum (ES) and the range where HS images are commonly captured depending on the sensor type. Charge-coupled device (CCD) silicon-based sensors are sensitive in the visible and very near-infrared (VNIR) spectrum in the range between 400 and 1000 nm. The standard indium gallium arsenide (InGaAs) sensors are able to capture HS images in the near-infrared (NIR) range, between 900 and 1700 nm, extending in some cases the upper range to 2500 nm. Other types of sensors can reach larger spectral ranges. For example, the mercury cadmium telluride (MCT) sensors are able to acquire HS images in the short-wavelength infrared (SWIR) range, from 1000 to 2500 nm, being able also to reach 25,000 nm in some specific systems [64].

HS cameras are mainly classified into four different types (Figure 2-3) depending on the methods employed to obtain the HS cube: whiskbroom (point-scanning) cameras, pushbroom (line-scanning) cameras, cameras based on spectral scanning (area-scanning or plane-scanning), and snapshot (single shot) cameras [50].

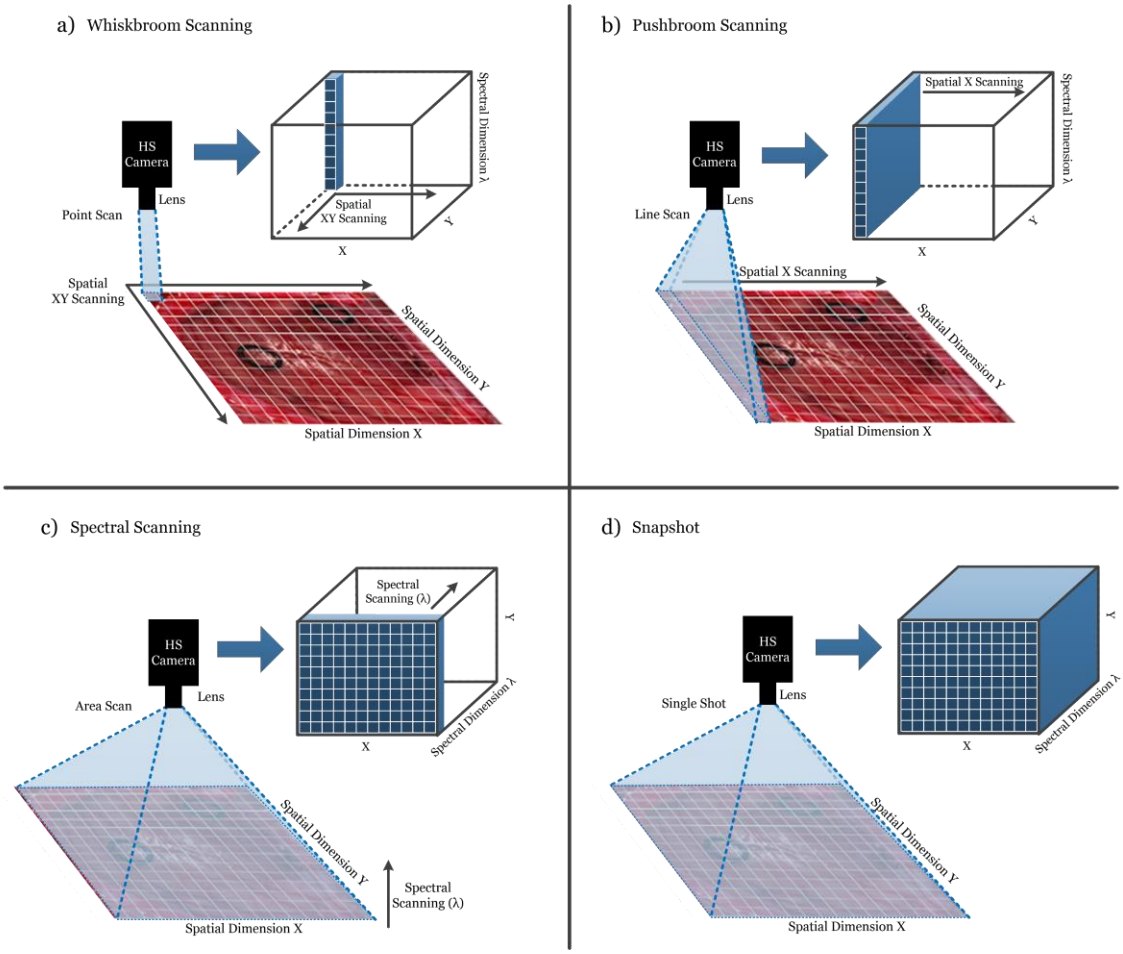

Figure 2-3. Hyperspectral camera types and their respective acquisition and data storage methods. (a) Whiskbroom camera; (b) Pushbroom camera; (c) Hyperspectral (HS) camera based on spectral scanning; (d) Snapshot camera.

Whiskbroom cameras (Figure 2-3a) are characterized by capturing one single pixel at one time containing all its spectral information. The rest of the pixels of the scene are captured by scanning both spatial dimensions (x and y). Whiskbroom cameras have the main disadvantage of being very time-consuming during the image acquisition. However, they can achieve very high spectral resolutions, allowing researchers to perform a more detailed analysis of the spectral signature of the pixel that is captured.

Closely related to the previous camera type, pushbroom cameras (Figure 2-3b) offer a faster scanning solution compared to whiskbroom, but also obtain high spectral resolution. In this case, the camera captures one line of pixels of the scene (y-axis) at one time. The other spatial dimension (x-axis) is obtained by moving the field-of-view (FOV) of the camera in that direction. Pushbroom cameras are the most common systems in remote sensing field and other industrial sectors due to their high spectral resolution, showing reduced capturing times with respect to the whiskbroom cameras. The main disadvantage of spatial scanning cameras is the presence of spatial aliasing if movement artifacts occur during the acquisition period [64].

On the other hand, HS cameras based on spectral scanning (Figure 2-3c) are able to obtain the entire spatial information (x-axis and y-axis) of the scene for a certain wavelength at one time, performing a scanning in the spectral dimension (λ-axis). These cameras can achieve high spatial resolutions and fast acquisition times; however, the spectral resolution is typically lower when compared to the spatial scanning (whiskbroom and pushbroom) cameras. One of the main disadvantages of spectral scanning cameras is that they are not suitable for capturing moving objects due to the time required to perform the spectral scanning, which would induce spectral aliasing artifacts. LCTF (Liquid Crystal Tunable Filter) and AOTF (Acousto-Optic Tunable Filter) HS cameras employ optical filters electronically controlled to filter each wavelength and generate the complete HS cube by performing a spectral scanning.

Finally, there is an emerging type of HS camera that can provide hyperspectral video, having the lowest acquisition time and allowing the acquisition of moving objects without performing any spatial or spectral scanning. Snapshot cameras (Figure 2-3d)

capture the entire scene in a single shot that contains both the spectral and spatial information [65]. The main disadvantage of snapshot cameras is that the spectral and spatial resolutions are much lower with respect to the other camera types.

In summary, the order of spectral resolution from highest to lowest is line scanning and point scanning being about equal, followed by spectral scanning, and snapshot. The order of spatial resolution from highest to lowest is spectral scanning, line scanning and point scanning being about equal, and snapshot. The order of fastest to slowest acquisition times is snapshot, spectral scanning, line scanning, and point scanning. The use of high spectral resolution HS cameras for cancer analysis allows researchers to study in detail the optical properties of the tissues, identifying the most relevant spectral channels that are involved in a certain application. On the other hand, the use of high spatial resolution HS cameras permits the analysis of samples by combining the spectral and the morphological properties of the tissue. Finally, HS snapshot cameras are the most suitable option for real-time analysis situations, mainly when the analysis of in-vivo tissue is performed. In conclusion, the HS sensor type and acquisition system selected are highly application dependent.

2.2.2 Hyperspectral Image Processing Algorithms

An extensive literature is available on the classification of HS images [66]. Traditionally, HSI has been widely employed in the remote sensing field, so the majority of algorithms developed to classify HS images are related to this field [67]. However, more recently, HSI is progressively being used in other fields, such as drug analysis [68, 69], food quality inspection [70], or defense and security [71], among many others. That is why

the algorithms that were developed targeting remote sensing application have been adapted to classify different types of scenes.

Pixel-wise classification methods assume that each pixel is pure or a mix of pure pixels and can be assigned to a certain material based on its spectral information [72-74]. Pixel-wise classification algorithms can be divided into two types: supervised classifiers and unsupervised classifiers (also called clustering or segmentation algorithms). Furthermore, in recent years, the use of deep learning (DL) approaches to classify HS data has become increasingly common, achieving excellent results when compared with traditional machine learning (ML) algorithms [75]. When applied to medical HSI data, these algorithms face two main challenges: the high dimensionality and the limited number of samples. However, these challenges are not necessarily present in other HSI domains, but are more prevalent in medical HSI because of substantial inter-patient spectral variability.

The basis of supervised classification algorithms relies on training an algorithm on a set of spectral signatures with known class labels and using this trained model to assign new labels to unknown spectral signatures in a HS image. The training process of supervised algorithms must be performed with a library of spectral signatures where each type of signature has been identified with a certain membership class, with the goal that this library is sufficiently representative for generalization purposes. Moreover, regression-based, statistical ML methods based on linear discriminant analysis, decision trees, random forest (RF) [75, 76], artificial neural networks (ANNs) [77-79], and kernel-based methods have been widely used to classify HS images. In particular, there are several types of kernel-based regression methods in the literature [80], where the support vector machine

(SVM) classifier is the most commonly used algorithm. In the HSI field, SVMs provide good performance for classifying this type of data when a limited number of training samples is available [80]. Due to its strong theoretical foundations, good generalization capabilities, low sensitivity to the problem of dimensionality and the ability to find optimal solutions, SVMs are usually selected by many researchers over other traditional, regression-based ML classification paradigms for classifying HS images [14]. As a relevant example, a variant of the SVM classifier, called fuzzy SVM classifier, was employed in the development of an emotion recognition system based on facial expression images [81]. In the medical field, SVMs have been used to detect multiple sclerosis in subjects, employing stationary wavelet entropy to extract features from magnetic resonance images used as inputs of the SVM classifier [82]. Furthermore, the same technique combined with a directed acyclic graph method has been used to diagnose unilateral hearing loss in structural MRI [83], demonstrating that the SVM algorithm is a reliable candidate to work with a variety of medical image modalities. Other relevant algorithms employed in HS classification problems with high dimensionality are linear and partial regression methods, such as partial least squares (PLS) and linear discriminant analysis (LDA), which are suitable for multivariate linear regressions with a large number of predictors; however, these perform well with small sample size, but may have problems generalizing to larger patient datasets [14].

On the other hand, the goal of the unsupervised classifiers is to divide an image into a certain number of similar groups (also called clusters), where each group shares approximately the same spectral information and provides the correspondent cluster centroid [84]. Each cluster centroid represents a spectrum corresponding to a material in

the scene, while the membership functions provide the weights for these spectra. Unlike the supervised classifiers, unsupervised methods do not require a training process using labeled samples. For that reason, they cannot provide the identification of the class that each pixel belongs to, only relative clustering with no information about the material's nature. Although unsupervised clustering does not provide any discriminant features by itself, it could be used to delineate the boundaries of the different spectral regions presented in an HS image. Unsupervised algorithms, such as the K-means algorithm [85, 86] and the Iterative Self-Organizing Data Analysis (ISODATA) technique [87, 88], are the most common clustering algorithms employed in the literature using HS data [66]. Hierarchical clustering is an unsupervised method of cluster analysis that seeks to obtain a hierarchy of clusters [89]. Several hierarchical clustering algorithms have been employed to classify HS images, such as Hierarchical Matrix Factorization [90], Hierarchical K-Means [91], and Hierarchical Spherical K-Means [92]. Some works based on HS analysis for medical applications use unsupervised clustering as part of the classification algorithms, such as those for colon tissue cell classification [93] or laryngeal cancer detection [94].

Deep learning (DL) techniques have been used for many applications of remote sensing data analysis, such as image processing, pixel-wise classification, target detection, and high-level semantic feature extraction and scene understanding [95]. Computationally, DL generates predictive models that are formed by several stacked processing layers with 'neurons' that can activate with learned weights to discriminate different representations of data with multiple levels of abstraction. DL architectures can extract intricate features in large datasets through an iterative, error backpropagation approach that determines the gradients that lead to successful changes of internal parameters [96]. While conventional

machine learning techniques are limited in their ability to process data in its input form, DL methods can learn new mathematical representations from the input data required for detection or classification. These multiple levels of representation are obtained by non-linear modules that modify the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level, where very complex functions can be learned [96].

Many DL frameworks have been applied to HS images in the literature. Of particular interest, convolutional neural networks (CNNs) [97, 98] have been employed to process and classify HS remote sensing data, improving upon the results obtained with conventional SVM-based algorithms [99, 100]. CNNs have also been employed to extract high-level spatial features from HS data in a spectral–spatial feature extraction algorithm for HS image classification [101]. In the medical field, DL is emerging in recent years as a powerful tool in the field of translational bioinformatics, imaging, pervasive sensing, and medical informatics [102]. As an example, deep neural networks (DNNs) and CNNs have been employed to classify electrocardiogram signals [103-105], detect retinal vessels [106-108], classify colorectal polyps [109, 110], and cancer analysis [111, 112]. On the other hand, despite so much promise using DL techniques, the use of DL for medical HSI is recent [113] because of the large amounts of training data required, and currently there are not many medical HSI databases available.

2.3 Cancer Optical Properties

The measured optical spectra of biological tissues from 400 to 1000 nm cover the visible and NIR regions and can be broken down at the molecular level, which greatly

contributes to the reflectance values measured in certain ranges. The hemoglobin (Hb) absorption and reflectance spectra vary substantially between the oxygenated and deoxygenated states, and Hb is a major spectral contribution of biological tissue in the range of 450 to 600 nm [114]. Deoxygenated Hb shows a single absorbance peak around 560 nm, while oxygenated Hb shows two equal absorbance peaks around 540 and 580 nm [115]. The region of the NIR spectrum from 700 to 900 nm corresponds with the scattering dominant optical properties of collagen [116]. The NIR region is referred to as the scattering dominant region for biological tissues, mainly composed of fat, lipids, collagen, and water. The molecular contributions of absorbance at wavelengths in the typical HSI range are shown in Figure 2-4. The relationship of absorbance is inverse to reflectance measured by HSI systems. For a more detailed summary of the optical properties of biological tissue, the interested reader is directed to the canonical review by Jacques [117].

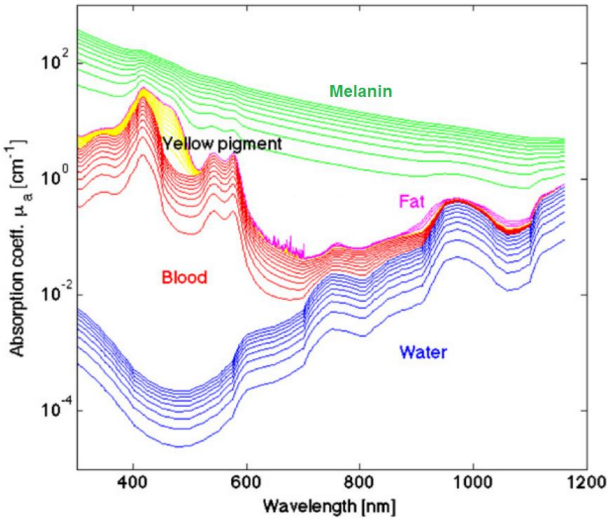

Figure 2-4. A few representative major molecular contributions to the absorbance at wavelengths of light typical for HSI investigations of biological tissue [117]. Reproduced with permission from [117]; published by IOP Publishing (2013).

Deal et al. investigated the contributions of normal and neoplastic colonic autofluorescence, which is the reflectance observed from the endogenous fluorescent molecules in biological tissue, using excitation wavelengths from 360 to 550 nm and emission at 555 nm, separated with a long-pass emission filter and dichroic beamsplitter [118]. Investigating neoplastic tissues, colorectal adenocarcinoma and adenomatous polyps, demonstrated that some autofluorescent molecules, such as elastin and nicotinamide adenine dinucleotide (NADH), had a significantly different abundance compared to normal colonic tissues, while other molecules, such as collagen, flavin adenine dinucleotide (FAD), and protoporphyrin IX (PPIX), showed no change between normal and neoplastic tissues. However, the authors acknowledge the limitation of performing their experiments with pairs of healthy and cancer from only nine patients and that non-neoplastic normal tissue from a diseased colon may vary molecularly from normal healthy colon tissue in a subject devoid pathology.

Monte-Carlo (MC) methods applied to medical HSI use simulations of photons with random parameter perturbations to simulate the interaction of light and biological tissues. Herrmann et al. developed an MC simulation of HS illumination in the visible and NIR regions to study reflectance signals of a multi-layer model in silico with different blood volume and oxygen saturations per layer [119]. The authors confirmed the absorbance peaks of water at 1000 nm and reflectance ratios at 580/800 nm for Hb. Interestingly, the authors also contend that blood volume fractions of 5% and 10% are detectable at depths of up to 1 mm in the simulated biological tissue, but changes beyond this depth are likely not resolvable. Additionally, it is possible to correlate in vivo measured reflectance values from various cancers and healthy tissues to extract meaningful optical

tissue properties associated with distinct molecular components, such as collagen, keratin, and Hb, using an inverse-MC approach, which has been previously reviewed for its application to cancer diagnosis [120].

2.4 Medical Hyperspectral Imaging for Cancer Analysis

In the previous sections, the basis of the HSI technology, the main algorithms employed to process this type of data and the optical properties of cancer tissue have been described. The following section is devoted to detail the state-of-the-art methods and primary research on the use of HSI within the medical field, focusing on cancer analysis.

The primary research performed in the literature related to the use of HSI for cancer analysis can be categorized first by organ systems, next the type of tissue samples and experimental design (ex vivo, in vivo), and finally the type of subjects (human, animal). Moreover, in vitro studies will not be included in this review of state-of-the-art works for cancer detection.

While it is possible to detail the investigations of HSI for cancer analysis altogether, HSI systems are not standardized, as different technologies were used in the following studies. Indeed, most of the studies work in the VNIR spectral region, employing CCD sensors. However, in some studies, the NIR region is also explored, requiring the use of InGaAs sensors. Halogen or xenon lamps are generally used as illumination systems for HSI applications, and sometimes optical fibers are used for light transmission to avoid the high temperatures produced by these types of light sources or to concentrate the light into a certain area. The main characteristics of the systems and analysis methods employed in

each study are presented in this literature review, sorted by the year of publication within each section.

2.5 Head and Neck Cancer

2.5.1 Clinical Need for HSI of Head and Neck Cancer

Head and neck (H&N) cancers are the sixth most common cancer worldwide [139]. Approximately 90% of the cancers at origin sites of the upper aerodigestive tract, which includes the oral cavity, nasal cavity, pharynx, and larynx, are squamous cell carcinoma (SCC). There are several well studied risk factors for H&N SCC, including consumption of tobacco and alcohol and oral infection with human papilloma virus (HPV) [140]. Patients with SCC typically present it at an advanced stage (stage 3 or 4 disease) [141]. The mainstay treatment for SCC is surgical cancer resection. The single largest predictor of patient outcomes for SCC resection is the successful removal of the entire SCC from the surgical wound bed, referred to as negative margins. The presence of positive or close (less than 2 mm of tissue clearance) margins after surgery greatly increases the likelihood for locoregional disease recurrence and additional surgeries [142]. Surgeons rely on intra-operative pathologist consultations with the surgical pathology department to ensure that negative margins are obtained through the use of frozen-section (FS) microscopic analysis of the resected specimens. Despite this, in the literature it is reported that up to 20% of patients will have a final diagnosis of positive or close margins despite having negative FS intraoperatively. Because of the difficulties in treating this challenging form of cancer, the estimated five-year survival rate of SCC is only 40 to 60% with treatment [143]. There

exists a great need to provide more near-real-time information and guidance to the operating head and neck surgeon.

2.5.2 In-Vivo Animal Head and Neck Cancer

Head and neck SCC was studied in-vivo using HS images from mice with SCC. The studies were performed in the VNIR range between 450 and 950 nm using a CRI system (Perkin Elmer Inc., Waltham, Massachusetts), which is comprised of a Xenon illumination source, LCTF, and a 16-bit CCD camera capturing images at a resolution of 1040 by 1392 pixels and a spatial resolution of 25 μm per pixel. Lu et al. published several works in this field, where the tensor decomposition, PCA and KNN methods were employed to perform a feature extraction and automatic classification, achieving a sensitivity of 94% and a specificity of 91% in the discrimination of tumor and normal tissue [30]. Furthermore, the group also studied the tumor margin during the surgical procedures performing an in-vivo/in-vitro registration between the in-vivo HS images and the histological images to validate the results [144].

On the other hand, their research has analyzed which pre-processing techniques are more suitable to compensate the variations of the environmental conditions during the acquisition inside an operating theatre [31, 145]. In the work published in 2015, a method based on the mRMR (maximal Relevance and Minimal Redundancy) algorithm was proposed to address the problem of glare that usually appears in the HS images, improving the sensitivity and specificity results to 94% and 98%, respectively. In addition, other techniques were studied, such as the use of a minimum-spanning forest (MSF) algorithm for an automatic classification and segmentation of the in-vivo HS images [146]. Figure

2-5 shows the results from the MSF algorithm applied to an HS image of an in-vivo mouse xenografted with a line of human head and neck SCC, obtaining an accurate identification of the head and neck tumor with respect to the gold standard.

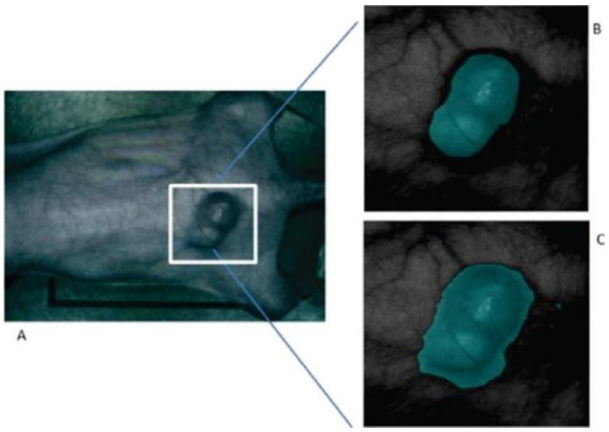

Figure 2-5. Result of the tumor identification using the Minimum-Spanning Forest method developed in [146]. (a) Synthetic RGB image of the original mouse; (b) Corresponding gold standard image; (c) Classification result obtained. Reproduced with permission from [146]; published by IEEE (2015).

2.5.3 Ex-Vivo Human Head and Neck Cancer

Active research into the application of HSI for H&N cancers is led by our group under professor Baowei Fei. Currently, all experiments explore H&N cancers including SCC and thyroid cancer in ex-vivo surgical tissue specimens and use the previously described CRI Maestro HS acquisition system in the VNIR spectral range, from 450 to 950 nm, with a spatial resolution of 1392×1040 pixels (25 μm per pixel), capturing 91 spectral bands.

In 2017, several works were published in this area with the goal of discriminating cancerous and non-cancerous tissue. Fei et al. achieved an accurate delineation of the boundaries between the normal and cancerous tissue using head and neck ex-vivo samples

compared with the histopathological results (Figure 2-6) [147]. The ensemble linear discriminant analysis (LDA) was employed to perform the classification, achieving an average accuracy, sensitivity and specificity of 90%, 89% and 91%, respectively, using oral cavity samples and an average accuracy, sensitivity and specificity of 94%, 94% and 95%, respectively, using thyroid samples. Autofluorescence, fluorescence with 2-deoxy-2-[(7-nitro-2,1,3-benzoxadiazol-4-yl)amino]-D-glucose (2-NBDG) and proflavine images were also classified and compared with the HSI results, demonstrating that HSI offered better results over the other alternative reflectance-based imaging modalities (an improvement of more than 7% of accuracy).

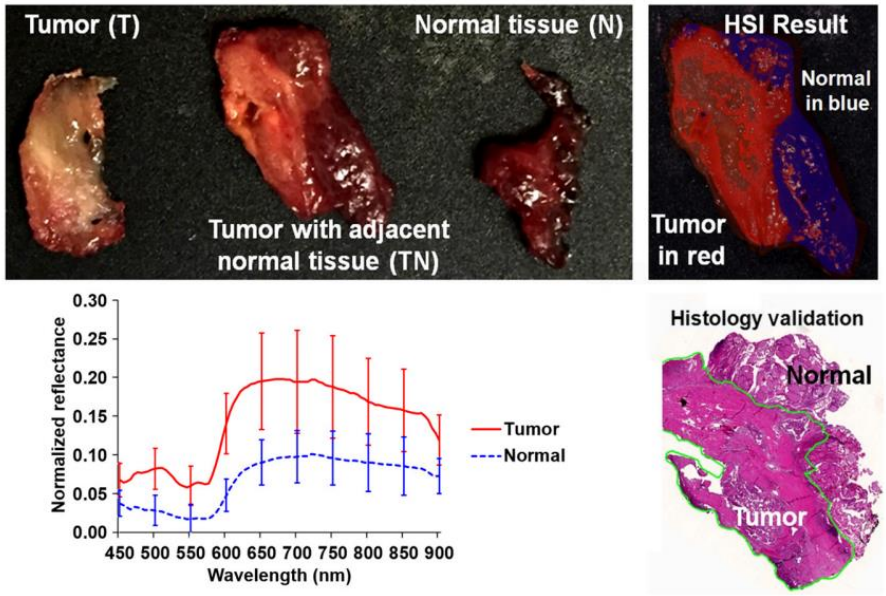

Figure 2-6. Preliminary results obtained in the tumor margin delineation for head and neck cancer [147]. After hyperspectral image acquisitions (top-left), the tissue was processed histologically, and tumor margins were outlined on the pathology image (bottom right) by a pathologist, which was used to validate the results of the classification (top-right). The average spectral curves are shown at the bottom left for each type of tissue, i.e., tumor, normal, and tumor with adjacent normal tissue. Reproduced from [147]; Creative Commons BY 4.0; published by SPIE (2017).

In addition, Lu et al. increased the number of patients (N=25) and performed a comparison using different machine learning classification approaches, reinforcing the conclusion obtained from the other study, where the ensemble LDA outperformed other traditional machine learning algorithms [33]. In this study, both intra-patient and inter-patient classifications were performed, as well as different classifications using different spectral regions within the VNIR range (450–600 nm, 605–850 nm, 855–900 nm, and 450–900 nm). Finally, the authors contend that the use of the entire spectral range (from 450 to 900 nm) provides the best accuracy results.

Recently, one of the first studies performed in the HS literature regarding the use of deep learning methods to classify HS images with the goal of distinguishing cancerous and non-cancerous tissue was performed by Halicek et al. [113]. The authors developed a CNN classifier to process the ex-vivo tissues from 50 different patients and compared the deep learning method with traditional machine learning approaches, demonstrating that CNNs outperform the traditional classifiers in this case. Additional works by Halicek et al. have not been summarized because the manuscripts are reproduced in full for this thesis in the following chapters and discussed in detail.

Lastly, several works have investigated NIR HSI for SCC detection in tongue specimens only, using leave-one-patient cross-validations in a dataset of 14 patients. Trajanovski et al. 2019 incorporated a U-net for semantic segmentation of SCC in 14 tongue specimens and achieved 0.93 validation AUC scores [148]. Using the same 14 specimens of tongue SCC, Brouwer de Koning et al. 2019 found 84% sensitivity for visible (400 to 950 nm) and 77% sensitivity for NIR (950 to 1,700 nm) HSI; interestingly, combining visible HSI with NIR did not provide additional benefit [149].

2.5.4 In-Vivo Human Aerodigestive Tract Cancer

A few studies can be found in the literature using HSI to analyze in-vivo samples of human subjects. Mainly, the studies are related to the use of endoscopic systems attached to an HS camera. In 2011, Kester et al. developed a customized real-time snapshot HSI endoscope system based on an image mapping technique and light dispersion that is capable of operating at frame rates of 5.2 fps, obtaining HS cubes of 48 bands in the visible range between 450 and 650 nm, with a spatial resolution of 100 µm [23]. Using this system, they were able to capture in-vivo tissue, resolving a vasculature pattern of the lower lip while simultaneously detecting oxy-hemoglobin. Figure 2-7 shows an example of the spectral signatures obtained by the system and the developed acquisition system.

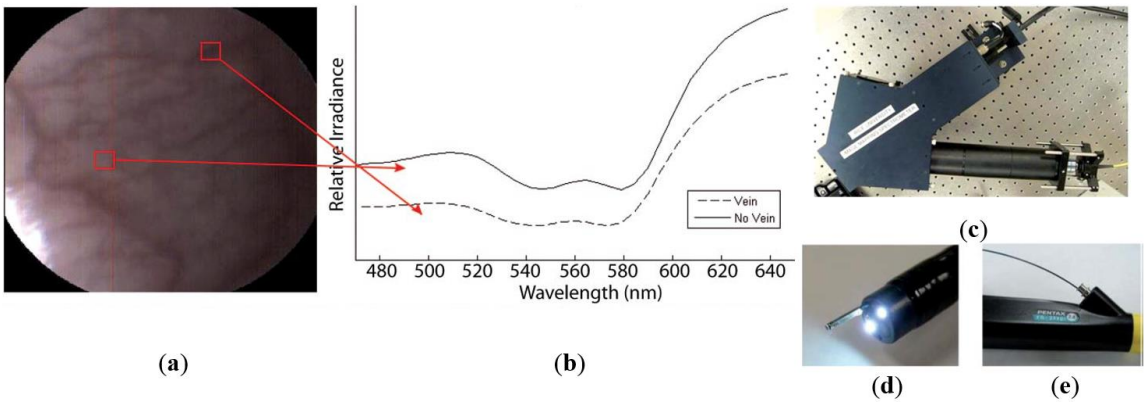

Figure 2-7. HS image example of the lower lip of a normal human acquired with the image mapping spectroscopy (IMS) endoscope developed in [23]. (a) RGB representation; (b) Spectral signature of the normal tissue pixel and a vein pixel; (c) Clinical setup of the IMS endoscope; (d) Miniature imaging end of the IMS endoscope; (e) Fiber optics of the IMS endoscope inserted into the instrument channel. Reproduced from [23]; Creative Commons BY 4.0; published by SPIE (2011).

Another study was published by Jayanthi et al. related to the use of diffuse reflectance spectroscopy for early detection of malignant changes in the oral cavity [150]. The system was able to capture HS information within the visible spectral range (from 400

to 700 nm) based on a snapshot light dispersion technique, obtaining 40 different bands. They used PCA for dimensionality reduction and LDA for automatic classification of the data. They achieved sensitivity and specificity results higher than 95% in the discrimination between different lesions, such as normal/healthy, hyperplastic, dysplastic and SCC tissues.

In 2016, laryngeal cancer was investigated by Regeling et al. using a flexible endoscopy coupled to an HSI system that was able to obtain HS images composed of 30 bands in the visual spectrum between 390 and 680 nm [94]. This system was employed to obtain in-vivo HS images that required substantial image pre-processing, such as registration due to the patient's heartbeat and noise removal due to specular reflections [151]. The images were registered using a rigid image-to-image registration based on normalized cross-correlation; the noise was reduced using the minimum noise fraction transformation; and the glare was detected using a customized method. For classification, a random forest (RF) algorithm was applied to distinguish between healthy and cancerous tissues, achieving an overall accuracy of 88%.

Also in 2016, Laffers et al. employed a rigid HS endoscopic system to capture HS images between 390 and 680 nm of the oral cavity and oropharynx from 85 patients [152]. However, in this study they only took into consideration three patients, one of them used for training the algorithm and the other two for validation purposes. The classification results obtained using the RF algorithm were sensitivities of 61% and 43%, and specificity of 100% in the two validation patients. These reduced sensitivity values could be mainly produced by the low number of patients involved in the training of the classification algorithm, which would not correctly handle inter-patient variability for the training phase.

Finally, tongue cancer of in-vivo human samples was studied in 2012 by Liu et al. using HSI [153] (Figure 2-8). The HS system utilized was based on an acousto-optic tunable filter (AOTF), capturing 81 bands in the VNIR spectral range comprised between 600 and 1000 nm. They developed a classifier based on the sparse representation method and compared the results obtained using traditional machine learning algorithms such as SVM and RVM (Relevance Vector Machine) classifiers. Sensitivity and specificity results of 91% and 94%, respectively, were obtained, increasing the accuracy by more than 4% with respect to the other two methods.

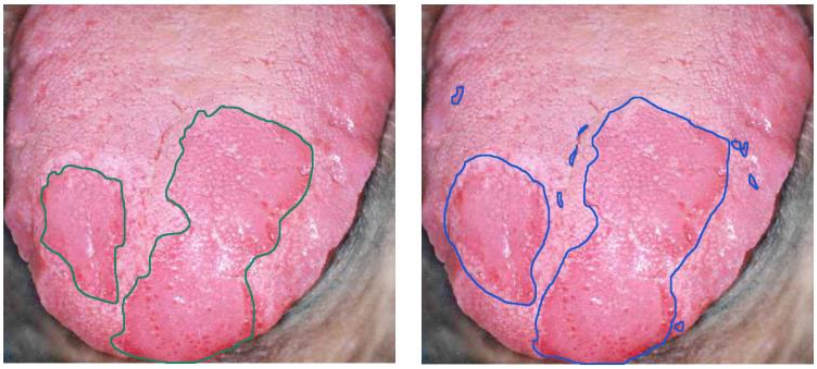

Figure 2-8. Delineation of the tongue tumor region in [153]. Expert labeling (left) and classifier prediction of tumor regions (right). Reproduced from [153]; Creative Commons BY 4.0; published by MDPI (2012).

2.6 Cancer of Other Organ Systems

2.6.1 Gastrointestinal Cancer

2.6.1.1 Clinical Need for HSI of Gastrointestinal Cancer

There are strong indications for endoscopy of the upper digestive tract and colonoscopy of the bowel for early detection of digestive tract cancers, with regular screenings for higher risk individuals [121]. Gastric carcinomas, most frequently caused by infection with helicobacter pylori or dietary contributions, are the second leading cause of cancer death globally. For the approximately two-thirds of gastric cancer patients that are diagnosed with locally advanced or metastatic disease, with a five-year survival of only 10%, the only curative therapy remains surgical resection [122]. Colorectal cancers, mostly adenocarcinomas, are thought to arise mainly from secondary risk factors of dietary and lifestyle origins, such as excessive caloric and fat intake, smoking, and physical inactivity [123]. Laparoscopic surgeries, performed in a way that is minimally invasive and guided by endoscopy, have been shown to be clinically equivalent in randomized controlled trials [124]. However, during this minimally invasive surgery, there is a loss of tactile feedback that surgeons often require, so there is a need to overcome this lost information [125]. HS imaging has been proposed as a solution to this problem with potential for more accurate digestive tract cancer resections.

2.6.1.2 Ex-Vivo Human Gastric Cancer

In 2011, Akbari et al. performed a study to identify gastric tumors in human ex-vivo tissues using an HS system, which was capable of capturing images in the range between 1000 and 2500 nm, obtaining 239 spectral bands [126]. An integral filter and the normalized cancer index (NDCI) was applied to perform an automatic classification of the tumor tissue determining the boundaries between tumor and normal tissue using histopathological analysis to validate their results (Figure 2-9). From their experiments, they determined that the spectral regions between 1226 and 1251 nm and 1288 and 1370

nm are the most salient ranges for distinguishing between non-cancerous and cancerous gastric tissue.

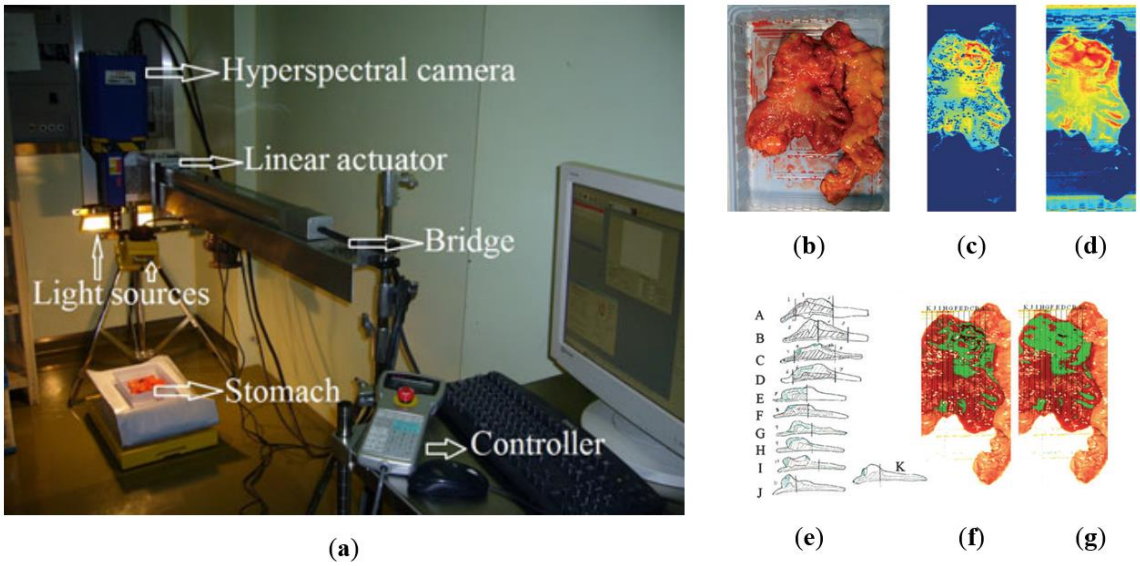

Figure 2-9. Gastric cancer detection acquisition system, cancer detection results using the NDCI and integral filter, and comparison with histopathological results obtained in [126]. (a) HS acquisition system setup; (b) RGB representation of the ex-vivo sample; (c) Cancer enhanced regions using an integral filter in the hyperspectral image (1057–2440 nm); the tissues are shown in a blue to red spectrum, where the red regions represent the tumor; (d) Cancer enhanced regions using NDCI; (e) Pathological sectioning and results; (f) Detected tumor using an integral filter; (g) Detected tumor using NDCI. Reproduced with permission from [126]; published by Wiley (2011).

In 2013, Kiyotoki et al. collected HS images in the spectral range comprised between 400 and 800 nm from ex-vivo tissue gastric samples to perform a preliminary study of gastroduodenal tumors removed by endoscopic resection or surgery from 14 different patients [127]. The system was able to obtain HS images comprised of 72 spectral bands with a spatial dimension of 640×480 pixels. Using these images, they were able to determine the optimal wavelength that allowed the most accurate classification between tumor and normal mucosa using the cutoff point method at the 726 nm wavelength. The sensitivity, specificity, and accuracy obtained in the test samples were 79%, 92%, and 86%,

respectively. This work was expanded upon in 2015 by the same group, increasing the number of patients to 96 and performing the selection of the optimal wavelength using the Mahalanobis distance, which in this case was 770 nm [128]. Sensitivity, specificity, and accuracy results obtained were 71%, 98%, and 85%, respectively, demonstrating that the increment in the number of patients to analyze did not decrease the accuracy of the method. Although the classification method employed to distinguish the different types of samples was quite basic, the studies revealed promising results in the use of HSI as a diagnostic tool for gastric cancer.

Baltussen et al. performed a study of laparoscopic HS imaging using two HS cameras collectively sensing between 400 and 1700 nm to distinguish normal fatty tissue, healthy colorectal mucosa, and adenocarcinoma in order to provide more diagnostic information back to surgeons, given the loss of tactile feedback during endoscopic procedures [129]. The authors utilized 32 patient samples to perform a three-class detection using quadratic SVMs of fat, muscle, and tumor, and obtained a tissue-level accuracy of 88% and a patient-level accuracy of 93%. One limitation of the presented HSI technique was focusing only on muscle and fat as normal tissues, but the authors acknowledge that future work should involve the entire specimen.

2.6.1.3 In-Vivo Human Colon Cancer

HS endoscopic systems have been used to study in-vivo colorectal tumors in the literature. One of the main studies in this field was performed in 2016, when Han et al. used a flexible hyperspectral colonoscopy system based on a motorized filter wheel, capable of obtaining 27 different bands in the range comprised between 405 and 665 nm,

to discriminate between malignant colorectal tumors and normal colonic mucosa in human patients [130]. They used a wavelength selection algorithm based on the recursive divergence method to identify the most relevant wavelengths in the spectral range employed, demonstrating that HSI can be used in-vivo for outlining the disease region and enhancing the microvascular network on the mucosa surface.

2.6.2 Breast Cancer

2.6.2.1 Clinical Need for HSI of Breast Cancer

Breast conserving surgery, also known as lumpectomy, with adjuvant radiation therapy, is the recommended surgical approach over traditional mastectomy without radiation for women diagnosed with early breast cancer. Women diagnosed with stage I or II breast cancer showed increased overall survival and disease-survival rates when treated with lumpectomy and radiation compared to complete mastectomy without radiation therapy, and this result was seen for all age groups and cancer types [131]. Successful breast conserving surgery is directly dependent on complete removal of the tumor mass with adequate margins, meaning there is a buffer of healthy tissue on the free cut edge. There is evidence to suggest that conservatively around 20% of women who undergo partial mastectomy have a final, post-operative diagnosis of positive margin status, which requires additional surgeries, with some studies reporting higher figures [132]. Intraoperative biopsies along with pathologist consultations are necessary tools to guide surgeons, but the need remains to provide more intraoperative diagnostic information with one potential solution being HSI.

2.6.2.2 In-Vivo Animal Breast Cancer

One of the first and most relevant works performed using HSI to study breast cancer was performed in 2007 by Panasyuk et al. [133]. In this work, a HS system based on LCTFs was used to acquire HS images in the visual spectral range, between 450 and 700 nm and composed of 34 bands, during intraoperative surgery of 60 rats affected by an induced breast cancer. They generated classification maps, where different types of tissue including tumor, blood vessels, muscle, and connective tissue were clearly identified. Furthermore, comparison to the histopathological examination of the tumor bed yielded a sensitivity of 89% and a specificity of 94% for the detection of the residual tumor by HS imaging. One of the limitations of this work was the use of light-emitting diode (LED) illumination in the HS acquisition system, which produced a non-standard spectral signature because LED light does not provide a uniform, broadband spectrum, such as that obtained by halogen or xenon light sources.

McCormack et al. performed a study of mouse models of breast cancer that aimed to evaluate the use of in-vivo HSI for microvessel oxygen saturation (sO2) monitoring during surgical procedures, studying also the response of the microvessels to different types of treatments [134]. The HS acquisition system was based on LCTFs and a halogen lamp, capturing images in the spectral range between 500 and 600 nm and composed of 26 bands because the absorption levels of both oxy and deoxy-hemoglobin are known to peak in this range.

2.6.2.3 Ex-Vivo Human Breast Cancer

Breast cancer has also been studied using ex-vivo samples with the goal of automatically delineating the regions of interest (ROI) in the samples and classifying the

tumor and normal tissue samples. In 2013, two studies were published with both previously mentioned goals using an HS system capable of obtaining images in the spectral range between 380 and 780 nm. The study conducted by Kim et al. performed an automatic ROI detection based on contrast and texture information achieving a true positive rate and a true negative rate of 97% and 96%, respectively, similar to the results obtained in a manual segmentation (99% and 96%) [135]. In the study performed by Pourreza-Shahri et al., the authors performed a feature extraction (using the Fourier coefficient selection features method) and a dimensional reduction (using the Minimum Redundancy Maximum Relevance method) to the HS images and then performed an automatic classification, using the SVM classifier with the radial basis function (RBF) kernel, of the tissue samples, differentiating between cancerous and non-cancerous tissue [136]. Sensitivity and specificity results of 98% and 99%, respectively, were obtained, demonstrating that HSI is a powerful imaging modality that has potential for use in the aided diagnosis of breast cancer. Lastly, Kho et al. utilized a U-Net CNN to perform breast cancer detection and segmentation with two HS cameras, one in the visible and one in the NIR regions, which achieved successful results when combining both spectra [137, 138].

2.6.3 Brain Cancer

2.6.3.1 Clinical Need for HSI of Brain Cancers

Brain tumors are categorized based on their histology and molecular parameters [154], with malignant gliomas the prevailing form of primary brain tumors in adults, causing between 2 and 3% of cancer deaths worldwide [155]. Surgery is one of the major

treatment options for brain tumors alongside radiotherapy and chemotherapy. Nevertheless, the surgeon's naked eye is often unable to accurately distinguish between tumor and normal brain tissue as brain tumors infiltrate and diffuse into the surrounding normal brain. During neurosurgeries, it is frequent that too much normal brain tissue is taken out (called safety margin) or that tumor tissue is unintentionally left behind (called residual tumor). Several studies have demonstrated that the residual tumor is the most common cause of tumor recurrence, and it is a major cause of morbidity and mortality [156, 157]. In contrast, over-resection of brain tumor tissue has been shown to cause permanent neurological damages that affect patients' quality of life [158].

Several image guidance tools, such as intraoperative neuronavigation, intraoperative magnetic resonance imaging (MRI), intraoperative ultrasound (US) and fluorescent tumor markers (for example 5-aminolevulinic acid, 5-ALA), are commonly used to assist surgeons in the delineation of brain tumors. Conversely, these technologies present several limitations. Intraoperative neuro-navigation is affected by the brain shift phenomenon [159], where the preoperative image link to the patient is affected by the brain deformation produced after craniotomy and durotomy. MRI significantly extends the duration of the surgery (between 20 and 75 min per image), generating a limited number of images and requiring special operating rooms [160]. On the other hand, US is inexpensive, real-time and unaffected by brain shift [161-163]. However, the use of US can lead to resection of histologically normal parenchyma. Finally, although 5-ALA is able to identify the tumor boundaries, it produces relevant side effects on the patient and should only be used for high-grade tumors [164, 165]. Thus, HSI can be a potential solution to

intraoperative margin delineation of brain tumors, being a label-free and non-ionizing imaging modality.

2.6.3.2 In-Vivo Human Brain Cancer

The goal providing intraoperative detection and delineation of brain tumors has been investigated by the European project HELICoiD (HypErspectraL Imaging Cancer Detection). Several studies have been performed to develop the HSI technology and machine/deep learning algorithms for in-vivo identification of the brain tumor margins [15, 166-169].

The main goal of which was to use HSI to generalize a methodology to discriminate between normal and malignant tissues in real time during neurosurgical procedures [168]. For this purpose, an intraoperative demonstrator was designed and built with the aim of acquiring intraoperative HS images and processing them in real time to assist neurosurgeons during resection [167]. This demonstrator captured HS images in the spectral range comprised between 400 and 1700 nm in approximately 2 min using two pushbroom cameras, one in the VNIR range (400 to 1000 nm) formed by 826 spectral bands with a high spatial resolution of 1004×1787 pixels and another HS camera in the NIR range (900 to 1700 nm) formed by 172 spectral bands with a low spatial resolution of 320×479 pixels. The HELICoiD brain cancer database is publicly available and consists of labeled samples from 33 HS cubes from 22 different patients [168]. Ravi et al. [15] investigated dimensional reduction approaches to evaluate and compare with a proposed modification of T-distributed Stochastic Neighbors (t-SNE). These methods were employed to generate three-band images from the HS cubes in order to provide high

contrast images used as inputs of a semantic segmentation classifier. In addition, due to the high dimensionality of the VNIR data, a pre-processing chain was proposed to reduce HS cubes to 129 spectral bands, which demonstrated increased performance for SVM-based methods [169].

In 2018, Fabelo et al. presented a classification algorithm for HS brain cancer detection using a chain of supervised classification methods combining SVM, PCA and kNN to combine spatial and spectral information, and the final result was obtained by majority voting with an unsupervised segmentation method, k-means, to delineate tumor boundaries [169]. The results were quantitatively and qualitatively evaluated following an intra-patient cross-validation method, achieving specificity and sensitivity results higher than 98%. Figure 2-10 shows the synthetic RGB representations of the HS cube and the corresponding thematic maps obtained when using this algorithm for normal brain and brain affected by cancer. In the thematic maps, the tumor tissue is represented in red color, the normal tissue in green, the hypervascularized tissue in blue and the background in black. The classification results obtained demonstrated the capabilities of HSI in the identification of different types of tumors, not only high-grade gliomas.

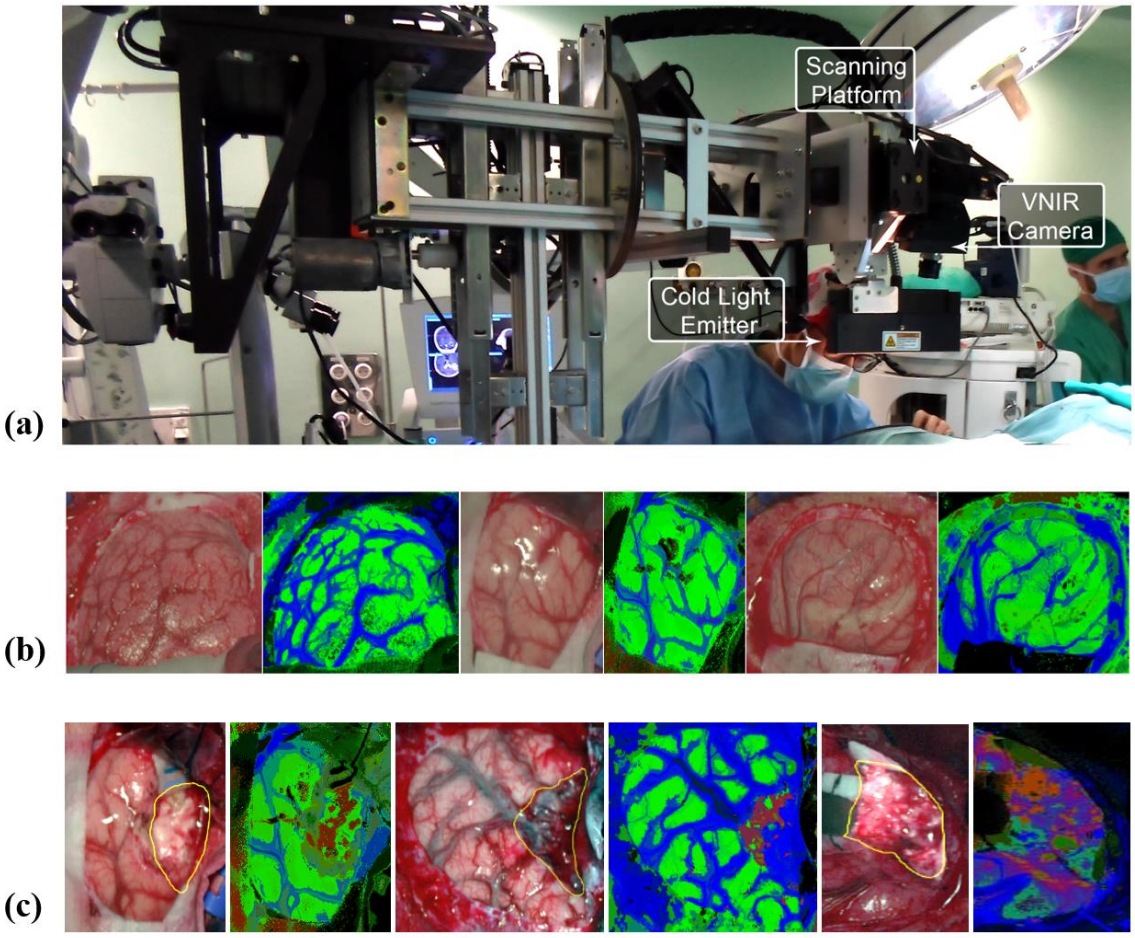

Figure 2-10. HELICoiD demonstrator set-up and normal and cancerous brain image results obtained from the database employed in [167]. (a) HELICoiD demonstrator. (b) Synthetic RGB images and classification maps of the normal brain HS images. (c) Synthetic RGB images and classification maps of the cancerous brain HS images, with the brain tumor outlined in yellow. For classification maps, predicted normal brain tissue (green), cancer (red), hypervascularized tissue (blue), and background (black). Reproduced from [167]; Creative Commons BY 4.0; published by MDPI (2018).

Lastly, in Fabelo et a. 2019, one of the latest studies related to the use of HSI for brain cancer presented a comparison between the use of SVM-based algorithms and deep learning approaches [166]. These experiments were carried out using only the glioblastoma tumor samples available in the database, 26 HS images from 16 patients, and performed

leave-one-patient-out cross-validation. The results obtained using deep learning architectures were promising, improving the accuracy of the tumor identification by ~16% with respect to the SVM-based algorithm results. However, these studies require a higher amount of data in order to validate the results obtained so far and also a clinical validation of the system should be carried out to assess it.

2.7 Discussion

Several questions that remain unanswered as we look back at hyperspectral imaging for cancer detection and analysis can be asked again. The best HS sensor for the clinical adoption of HSI necessary to make the leap from basic research to clinical translational medicine is unknown and debatable. The sensor is likely to be task specific. We have reviewed that line-scanning pushbroom HS acquisition systems produce higher spectral resolutions, but spectral scanning HS cameras allow for higher spatial resolution. For example, in clinical cases where real-time is not a critical issue but a higher spectral resolution is required, a pushbroom HS camera is appropriate, but in cases where there is need for faster acquisition and where fewer spectral bands are required, spectral scanning or snapshot HS cameras would work well. Traditional, regression-based algorithms may yield optimal performance and accuracy using only spectral signatures as inputs when the HS sensor exhibits high spectral resolution. However, there are certain tasks that also require interpretation of spatial information, and these types of tasks may require the use of a CNN or other method for contextual spatial information along with the spectral signatures.

Regardless of the sensors and HS acquisition system implemented, it is yet unknown which wavelength range of the electromagnetic spectrum is optimal for each application, and this could very well be again task specific. Most of the works reviewed in this article have been performed in the range of 400 to 1000 nm, broadly. Currently, as discussed in the HS sensor section, we appear to be lower-limited to 400 nm in the short wavelength range, but in the NIR and IR (infrared) range, it is possible to extend beyond 1000 nm, as some works have investigated. As reviewed in Section 2.3, the optical properties of biological tissue vary in spectra. It could be useful to have HS cameras below 400 nm for tasks sensitive to FAD (flavin adenine dinucleotide) or NADH. On the other hand, some tasks might require extending into the SWIR range for fat, water, or collagen analysis. Additionally, with regards to the broad-band spectrum, different algorithms that have been previously deployed and validated on other parts of the spectrum may not be generalizable in different ranges. It may happen that some algorithms work better compared with others for certain parts of the spectrum. For example, deep learning may have the potential to learn and tolerate more noise in the input signals, so a spectral range that contains more noise, either organic or systemic from the sensor, could be handled better by a CNN [140]. However, deep learning can be prone to over-fitting if there is insufficient data, and false negatives or positives can sometimes be predicted with exceptionally high confidence that the prediction is correct, which can lead to difficulties in interpreting the results. Therefore, it is conceivable that there are scenarios where traditional methods, such as SVM, may suppose the best choice for some specific works.

There has been a wide variety of works performed both ex-vivo and in-vivo, but there are challenges, both known and unknown, in moving from the former to the latter.

Mainly, it is unknown whether ex-vivo HS data correlates well with in-vivo data. If an algorithm is trained on ex-vivo data, there is no guarantee that the algorithm can be generalizable enough to apply to in-vivo data for testing or for clinical use. What is more, it cannot be assured that the same algorithms would work for the same task when moving from ex-vivo to in-vivo. For example, the in-vivo data collection could degrade the quality of the HS data or induce artifacts from the patient and surgical environment, and this could result in a certain family of algorithms being less effective than others. Therefore, if this phenomenon was observed, then entire datasets of large numbers of ex-vivo patient data need to be re-collected if the system is adopted for in-vivo clinical use.

Currently, several systematic limitations of general HSI hinder its use in the operating theater or in the clinic. The most obvious one from a user perspective will be the size of the machinery and apparatus. The imaging time, which can be up to one minute, can be another limitation. With the increased imaging time, there is more room for artifact induction due to reasons such as patient movement or instrumentation movement. In addition, there are several potential problems when imaging biological tissue. Because the light sources are so close to the HS sensors and the biological tissue is often wet, there is a problem with specular glare. The human body is not topographically uniform, so imaging a small area with changes in elevation can create inconsistent illumination of the scene or cast shadows. Aside from inconsistencies with illumination, this could lead to issues with the HS camera focusing on multiple imaging planes. It is difficult to forecast the effects of this on large-scale HSI, and this could lead to issues with reproducibility or false negative/positive results because of insufficient HS quality to produce an accurate and

reliable result. A level of quality control is necessary where the algorithms can automatically detect if the scene is not correctly captured to make an accurate prediction.

In conclusion, the reviewed studies present promising results for a wide variety of cancer detection applications based on medical imaging and for surgical guidance. However, currently we are limited in the field of HSI by technology restrictions and small datasets. The preliminary results warrant further research across all organ systems to determine if HSI has a place in the operating theater or in the clinic. Since HSI is non-contact, non-ionizing, non-invasive, and label-free, it is an attractive imaging modality with great potential. Most importantly, like all future translational technologies, HSI needs to be evaluated to demonstrate that it can be reliable, reproducible, and generalizable before it takes its place in medicine. The following chapters of this thesis work will explore HSI and deep learning on sufficiently large patient datasets, performing the crucial comparisons with RGB imaging and other forms of optical imaging.

CHAPTER 3. HYPERSPECTRAL IMAGING OF HEAD & NECK CANCER EX-VIVO TISSUE SPECIMENS: REGISTRATION, SPECULAR GLARE, AND VARIANCE OF THE TUMOR MARGIN

In this chapter, the necessary pre-processing and machine learning approaches for SCC detection using HSI is evaluated with a subset of *ex-vivo* surgical specimens. Additionally, the registration method and effectiveness of histological ground truths for gross-level *ex-vivo* specimens are developed and evaluated. Several machine learning methods are investigated, including convolutional neural networks (CNNs) and a spectral-spatial classification framework based on support vector machines (SVMs). Quantitative results demonstrate that additional data pre-processing and unsupervised segmentation can improve CNN results to achieve optimal performance when the training dataset is small. The methods are investigated in two paradigms: with and without specular glare. Classifying regions affected by specular glare degrade the overall performance, but the combination of the CNN probability maps and unsupervised segmentation using a majority voting method produces better results if the training data is limited. It is well known that different wavelengths of light used in HSI have different penetrate depths into biological tissue. Cancer margins may change with depth and create uncertainty in the ground-truth. Through serial histological sectioning, the variance in cancer-margin with depth is investigated and paired with qualitative classification results. Through compounding error of registration and depth variability, the validity of the ground truth may be limited to 1-2 mm. This chapter focuses on the study of specular glare, registration, pre-processing, and

margin variation to provide better understanding of the potential of HSI for use in the operating room.

3.1 Introduction

Head and neck cancer is the 6th most common cancer world-wide, with majority of cancers of the upper aerodigestive tract, comprising the oral and nasal cavities, pharynx and larynx being squamous cell carcinoma (SCC) [2]. Approximately two-thirds of SCC patients present with advanced disease, either stage III or IV [141]. Surgical resection is the primary management of SCC of the aerodigestive tract, potentially with concurrent chemo-radiation therapy depending on the extent of the disease [4]. The local recurrence rates for SCC cases depend on the successful removal of the cancer. For surgeries with negative cancer margins, the local recurrence rate is 12-18%, compared to local recurrence rates up to 30-55% if positive cancer margins are determined [8-10]. Moreover, positive cancer margins have a greatly reduced disease-free survival, with estimates ranging from 7 to 52%, compared to disease-free survival rates of 39-73% for negative margins [12, 13]. Disease recurrence greatly affects likelihood for additional surgeries, reduced quality of life, complications from surgery, and increased mortality rates [11].

Hyperspectral imaging (HSI) is a non-contact optical imaging modality capable of acquiring a single image of potentially hundreds of discrete wavelengths. Preliminary research demonstrates that HSI has potential for providing diagnostic information for various diseases [14]. Preliminary studies from our group show that HSI combined with machine learning may yield diagnostic information with potential applications for surgical use in head and neck cancers [33, 170-172]. Fabelo *et al.* have demonstrated the need and

utility of HSI for *in-vivo* brain cancer detection and developed a visualization system that could lead to near-real-time guidance by using and machine learning-based classification algorithm [166, 167, 169]. Moreover, the use of deep learning techniques to process the same dataset was study, demonstrating that they outperforms the traditional machine learning methods [166]. Many applications of HSI for gastrointestinal procedures have been explored, including anatomical organ identification, anastomosis and ischemia classification, and cancer detection [173]. Recently, the implementation of HSI for guided colorectal surgery as a replacement for tactile information lost in laparoscopic surgeries has been explored in an *ex-vivo* study, and Baltussen *et al.* demonstrated that HSI could distinguish between normal healthy colon wall and colorectal adenocarcinoma [129]. Much work has been explored in head and neck HSI as well. Salmivuori *et al.* detailed the assistance of HSI in ill-defined, cutaneous basal cell carcinomas of the head and neck, and the authors contend that HSI offers clinical utility in more accurately assessing margins [174]. Farah *et al.* demonstrated that narrow band imaging (NBI) at green (400–430 nm) and blue (525–555 nm) visible light could reveal the extent of oral SCC under white light [175]. Non-encoding portions of RNA called micro-RNA function to regulate the expression of DNA [176]. Moreover, Farah *et al.* obtained micro-RNA and mRNA expression levels from both the primary tumor core and near-tumor normal tissues; it was determined that NBI spatially correlated with tumor-like expression levels in detecting the abnormality in near-tumor normal tissues [177, 178].

Machine learning methods, including support vector machines (SVMs) and convolutional neural networks (CNNs), the latter being an implementation of artificial intelligence, have demonstrated near human-level ability for image classification tasks [26,

27]. These experiments have been conducted on gross-level HSI acquired of *ex-vivo* tissues specimens from patients undergoing surgical cancer resection. However, HSI uses wavelengths of light that can penetrate different depths into biological tissue, so it is possible that the superficial cancer margin may change with depth and create uncertainty in ground truth, which is used to obtained evaluation metrics for these experiments.

Selected previous works from our group performing head and neck SCC detection in *ex-vivo* samples from the upper-aerodigestive tract used only single class tissue specimens (purely tumor or normal tissues) to demonstrate classification potential with HSI [113, 179]. Lu et al. 2017 incorporated tumor-involved cancer margin specimens from head and neck SCC, but these methods were implemented on selected ROIs of normal and tumor near the margin [33]. Therefore, the AUCs reported for head and neck SCC detection thus far from our group ranges from 0.8 to 0.95, but these results may be limited to ROIs or single class tissues. Similarly, Manni et al. 2019 used ROIs from tumor-margin specimens from 7 patients with tongue SCC and obtained an AUC of 0.92 [180]. Weijtmans et al. 2019, developed a deep learning architecture that separately extracts spectral and spatial features from HSI, a dual-streamed approach, also validated on 7 patients with tongue SCC; the proposed model performed better with both feature streams (AUC of 0.90) compared to individually [181]. A few works have investigated SCC detection at the actual cancer margin; Halicek et al. 2018 performed SCC detection at the cancer margin, upon which the proposed method in this paper expands, but this previous work was limited by ignoring regions of specular glare [182]. Trajanovski et al. 2019 performed semantic segmentation with deep learning of entire *ex-vivo* tissue specimens

with excellent validation performance (AUC of 0.93) on gross specimens of the cancer margin from 14 tongue SCC samples [148].

This study aims to investigate the ability of HSI to detect SCC in surgical specimens from the upper aerodigestive tract using several distinct machine learning pipelines. Additionally, another objective of this work is to investigate the limiting factors of HSI-based SCC detection, including specular glare, noise and blur, and uncertainty in the ground truth due to changes in superficial cancer margin with depth, all of which must be thoroughly explored to understand the potential of HSI in the operating room. A preliminary version of this work was presented at the 2019 SPIE Medical Imaging Conference [182]. The new contributions of this journal paper include a proposed algorithm combining deep learning and unsupervised machine learning, additionally five testing patients have been serially sectioned to reveal changes in the cancer margin. This work expands upon previous cross-validation experiments with CNN-only methods on our H&N dataset to include multiple machine learning pipelines involving CNNs and other state-of-the-art methods. The proposed methods are tested on five HSI from five SCC patients, and the accuracy of the corresponding ground truths from these five tissues are discussed in detail along with potential directions for incorporating the outcomes of this work to improve future studies with HSI for cancer detection.

3.2 Methods

3.2.1 Experimental Design

In collaboration with the Otolaryngology Department and the Department of Pathology and Laboratory Medicine at Emory University Hospital Midtown, head and neck

cancer patients undergoing surgical cancer resection were recruited for our hyperspectral imaging studies. Written and informed consent was obtained from all patients before acquiring surgical tissue specimens for inclusion in our study, which was used for research purposes only and de-identified by a clinical research coordinator. The experimental methods and protocols were approved by the Institutional Review Board at Emory University. In previous studies, we have evaluated the efficacy of using HSI for optical biopsy of head and neck tissues [147, 170]. Excised tissue samples of head and neck squamous cell carcinoma (HNSCC) and normal tissue were collected from the upper aerodigestive tract sites, including tongue, larynx, pharynx, and mandible. Three tissue samples, with approximate size of 10×10×3 mm, were collected from each patient: a sample of the primary tumor specimen, a normal tissue sample, and a sample at the tumor-involved cancer margin with adjacent normal tissue; these specimens were scanned with a HSI system [33, 113].

In this study, we selected 26 tissue specimens from 12 patients with moderate to poorly differentiated primary SCC of the upper aerodigestive tract for this analysis, including primary origin sites of the larynx, pharynx, tongue, floor of mouth, alveolar ridge, buccal mucosa, and maxillary sinus. The patients were divided into two groups, cross-validation and testing groups. The first 7 patients and corresponding 21 tissues samples were collected early in the course of this project that met the criteria of having the ideal distribution of tumor, tumor-normal and normal tissues, and were used for evaluating quantitative results, expanding from previously published results with this group of patients [172]. Additionally, 5 tumor-involved margin tissue specimens from 5 patients acquired at the end of the data collection period were selected to comprise the testing group. These

tissue specimens were selected because they had approximately equal amounts of cancer and normal tissue in each specimen and were verified by the collaborating pathologist as of interest to undergo serial histological sectioning through the depth of the tissue. These 5 patients' tissues were classified using models trained from the first 7 patient cross-validation group (following a leave-one-patient-out cross-validation method), and these 5 patients' results are presented qualitatively to compare with the variability in the superficial cancer margin.

3.2.2 Hyperspectral Imaging

Hyperspectral images were acquired of *ex-vivo* surgical specimens using a previously described CRI Maestro imaging system (Perkin Elmer Inc., Waltham, Massachusetts), which captures images that are 1,040 by 1,392 pixels and a spatial resolution of 25 µm per pixel [113, 145, 170, 183]. Each hypercube contains 91 spectral bands, ranging from 450 to 900 nm with a 5 nm spectral sampling interval. The HS data were normalized by a standard white-dark calibration normalization, which involved subtracting the inherent dark current from the measured spectra and dividing by a white reference spectra for all wavelengths sampled for all pixels [145, 170]. Figure 3-1 shows the RGBs, grayscale images at representative bands, and spectral signatures of the patients' HSI data. The grayscale images at selected bands highlight the choice of the cutoffs due to noise at the ends of the broadband spectrum. The HSI was used to construct RGB composite images by implementing a Gaussian function centered on each color component of the spectrum: red (625 to 700 nm), green (520 to 560 nm), and blue (450 to 490 nm).

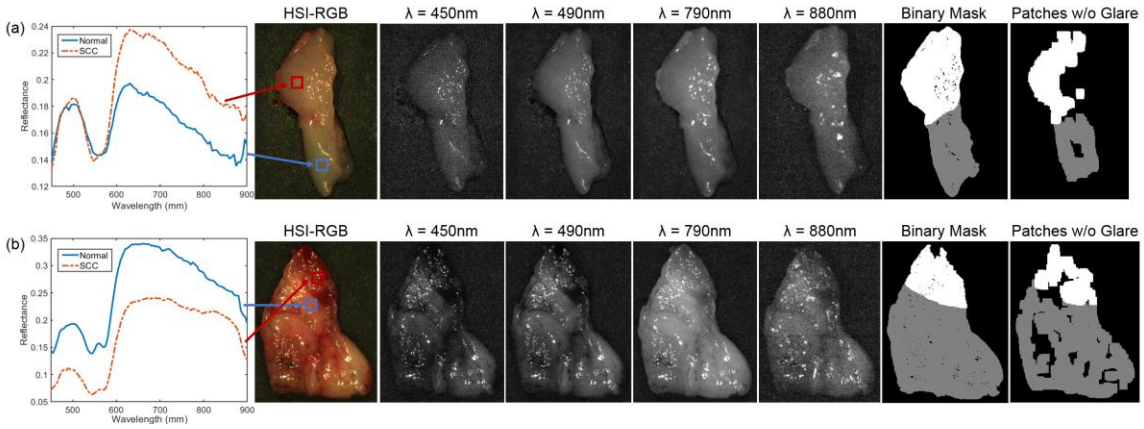

Figure 3-1. Two representative tissue specimens from different patients. Left: spectral signatures of SCC and normal ROIs are shown for both patients in (a) and (b). The HSI-RGB composite images and selected spectral bands are shown to highlight the noisy band cutoffs during pre-processing. Binary ground-truth masks including glare regions, generated by only removing patches centered on specular glare is shown; binary ground-truth mask excluding glare regions, generated by sufficient area to extract 25×25 patches and avoiding specular glare (white: SCC, and normal: grey).

3.2.3 Histological Imaging

To obtain a ground-truth labelling for the tissue specimens imaged with HSI, tissues were inked after imaging to preserve orientation, fixed in formalin, paraffin embedded, sectioned from the top of the imaging surface, haemotoxylin and eosin (H&E) stained, and digitized using a Hamamatsu Photonics NanoZoomer at 40x objective (specimen-level pixel size, 0.23 μm). The *ex-vivo* tissue sections were reviewed by a board-certified pathologist with expertise in H&N pathology, and the cancer margins were annotated directly on the slide using digitized histology in Aperio ImageScope (Leica Biosystems Inc, Buffalo Grove, IL, USA).

There exists the possibility for substantial deformation of the histological images during processing relative to the HSI setup. The registration challenges are explored using a deformation-based image registration pipeline of the histological image with known

registration landmarks, and an error of 0.5 mm for SCC tissues was calculated [184]. For the cross-validation group patients in this work, the histological images were automatically registered using this deformable registration pipeline to produce the HSI ground truth. However, to calculate the error of histology ground truths for HSI in millimeters, a more accurate registration is needed than the automated approach. For the testing group patients, since each had 6 histological slices, we manually registered each histological image using mutual landmarks to the HSI that were serially sectioned to reduce error.

3.2.4 Deformable Registration Pipeline

A binary mask is created from the HSI RGB composite image. A binary mask is also created from the histological image. These two binary masks are used as the input of the registration algorithm that was implemented with MATLAB (MathWorks Inc, Natick, MA, USA). A flowchart of the proposed registration method is shown in Figure 3-2. For samples with needle-bored holes, the average target registration error (TRE) was calculated from the center of the needle-bored hole after registration. An affine registration is performed using rotation, translation, and scale to produce a non-reflective similarity transformation [185]. This step is necessary to get the histology mask approximately aligned to the neighborhood of the HSI mask. If the affine registration is not sufficient by visual inspection, an optional control-point registration is applied using control-point pairs selected from the tissue edges of the masks. The control-point registration is implemented by a local weighted mean of inferred second degree polynomials from each neighboring control-point pair to create a transformation mapping [144, 186].

Demons registration is based on an analogy to Maxwell's demons, in which every pixel in an image is an effector with a force, referred to as a demon, in total used to create a deformation field. A free-form deformation is made from the individual displacements of the complete grid of effectors (demons) in an image, and a Gaussian filter is applied to smooth the image and to create a regular displacement field [187]. The displacement field is determined by points scattered along the contours of binary masks. This process is performed iteratively, and the displacement of each effector at each iteration is calculated using optical flow, which represents small displacements derived from diffusion model behavior [187, 188].

The demons registration method was applied in the deformable registration pipeline [187]. After the affine and control-point based registration, the transformed histology mask is registered to the HSI mask using deformable demons registration with an accumulated field smoothing value of 0.5 for five pyramid levels with one thousand iterations per pyramid level [187].

To evaluate the efficacy of the proposed registration methods, 17 tumor-margin tissue samples from 11 HNSCC and 6 thyroid cancer patients were registered. Dice similarity coefficient (DSC) was calculated by measuring the overlap of the histological and HSI mask. Target registration error (TRE) was calculated using both needle-bored holes in samples when available and tissue landmarks visible between gross and histology to obtain an average Euclidean distance representing error [144]. The TRE and DSC were calculated for each tissue sample, and averages were obtained for both groups.

Figure 3-2. Flowchart of the proposed registration method for histological cancer margins to HSI of surgical tissue specimens.

3.2.5 HSI Binary Ground-Truth

Specular glare is detected in each spectral band by fitting a gamma distribution to the pixel intensities of that band. Next, the top 1% of intensities are identified in this distribution, and any pixels with intensities of these bins are identified as specular glare. There were two types of specular glare observed: near-infrared (NIR) and visible, which usually were spatially independent. Visible glare occurs between 450 to 790 nm, and NIR glare occurs after 800 nm, which can be seen in the selected bands in Figure 3-1. A pixel was identified as a specular glare pixel and removed if there was glare identified using the distribution method in any spectral band from 490 to 790 nm. This range was selected because bands before and after this range were removed during pre-processing. NIR glare is observed beyond this range, so it is not relevant.

Binary masks are constructed from the histological images and used to construct the ground truth for HSI in two methods. The first method is to investigate only ideal quality pixels, which is constructed by avoiding HSI regions with a large amount of specular glare. For the ground-truth mask avoiding glare, only regions that contain sufficient area to extract 25×25 spatial patches without any specular glare are included. The second method is to investigate the degrading effect of specular glare on classification accuracy. For the ground-truth mask including glare regions, the entire tissue area is

included for patch-making, but the top 1% of glare pixels are identified by fitting a gamma distribution to pixel intensities, such that no patches are centered on a glare pixel. A binary mask of the specular glare pixels was generated, and patches were produced automatically using this mask in the regions around the specular glare. Figure 3-1 shows an example of the two masks generated from two different patients. Both masks will be used for evaluating quantitative testing results and are referred to as 'ground-truth mask excluding glare', and 'ground-truth mask including glare', respectively. The specular glare pixels are represented by the black pixels within the first mask in the middle. The automatic patch-making process can be seen to avoid the regions of specular glare shown in the second mask on the right of Figure 3-1.

3.2.6 Effect of Sectioning Depth on Cancer Margin Ground-Truth

HSI uses wavelengths of light in the visible and near-infrared spectrum that can penetrate different depths into biological tissue. It possible that the cancer margin may change with depth and create uncertainty in the classification results because of the variable penetration of the HSI broad-band spectrum. To investigate this, additional histological sectioning was performed on 5 tumor-involved margin tissue specimens (see Figure 3-3).

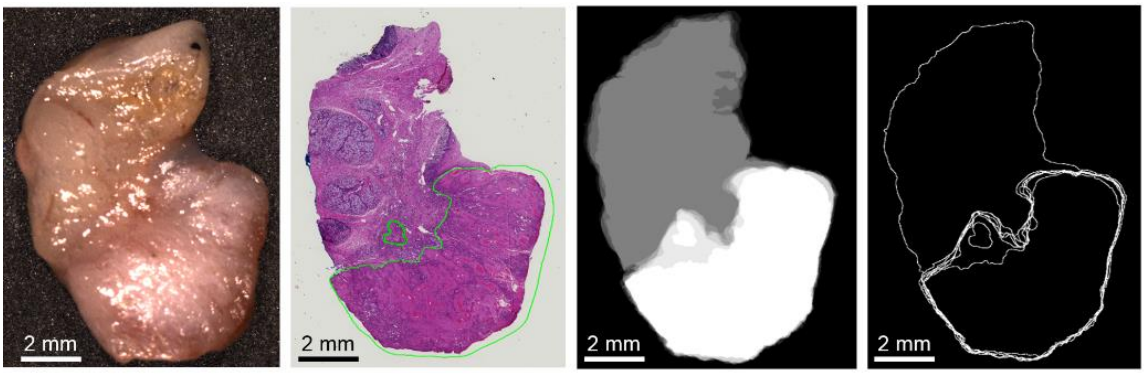

Figure 3-3. Representative tissue specimen that underwent serial histological sectioning to evaluate cancer margin variation with tissue depth. Left to right: HSI-RGB composite image; first histological slice (the green outlined area is the pathologist annotation of the SCC region); combined merged image of the 6 binary masks from the 6 histological images (cancer shown in white, normal in gray); merged image of the SCC contours from the 6 histological images. The two rightmost images are used to depict the total variation of the cancer margin.

During histological sectioning, it is standard procedure to discard the first sections until the first good quality section is obtained that contains the entire perimeter of the tissue specimen. The five specimens used for additional histological sectioning were retroactively enrolled for serial sectioning, so it was unknown what initial depth of tissue was discarded in the first sections before the first to obtain the first good quality slice. In discussion with the pathology laboratory technician, who performed the histological sectioning, and through observation of specimen sectioning, it was estimated that approximately 100-200 microns were discarded on average during initial sectioning before obtaining the first high-quality, tissue-encompassing slide.

To produce serial histological sectioning further into the depth of the tissue specimens, the thickness of the remaining paraffin embedded tissue was estimated. The microtome used for this study produced slices at 5 microns, so the number of additional sections was documented and the distances of additional depth could be measured. From

the sectioned surface, which corresponds to the top of the tissue that was optically imaged, five more sections were obtained, up to 300 microns further into the depth of the tissue, by discarding non-included slices. The extent of the additional serial sections were obtained for 100 to 300 microns beyond the first good quality slice; the exact values for the five tissue specimens' total additional sectioning depths were 100 microns, 150 microns, 200 microns, and 300 microns, depending on the unique tissue specimen. Therefore, the combined total estimate of additional sectioning is approximately 200 to 500 microns into the tissue depth from the original HS image surface. Penetration depth depends on the tissue composition and wavelength of light. This sectioning depth was investigated because it represents the effective penetration depth for biological tissue in the shorter wavelengths of HSI used, which correspond to the most relevant features for HSI classification, such as hemoglobin between 550 to 600 nm [189-191].

3.2.7 Machine Learning Techniques

3.2.7.1 Data Pre-processing

A further pre-processing chain was applied to the data mainly to reduce the noise in the spectral signatures (Figure 3-4) and compared with the standard white-reference calibration. The machine learning methods detailed below were tested with and without the following pre-processing steps. The proposed pre-processing chain is based on 4 steps: image calibration, operating bandwidth selection, noise filtering, and data normalization. In the first step, the previously described method to normalize the data using white reference and dark current is performed. Next, the spectral bands are truncated between 490-790 nm, so the final HS cube contains only 61 bands, which will be used for

classification. Spectral bands outside this range were noisy because they were too close to the detectable limits of the HS camera. In the third step, a smoothing filter is applied to the data in order to reduce the spectral noise. A moving average with a span of 10 spectral bands was used as the smoothing filter, implemented as a sliding low-pass filter with coefficients equal to the reciprocal of the span. Finally, each pixel's spectral signature is normalized between 0 and 1, using a function to map the minimum and maximum to these values.

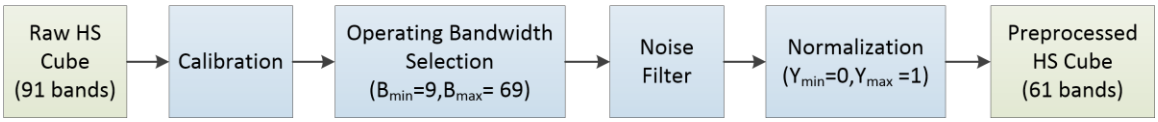

Figure 3-4. Block diagram of the proposed pre-processing chain.

3.2.7.2 Convolutional Neural Networks

A deep convolutional neural network was used to detect SCC in the *ex-vivo* specimens of the upper aerodigestive tract, implemented using TensorFlow in Python on a NVIDIA Titan-XP GPU [26, 27, 172, 192]. In summary, an Inception-V1-style 3D CNN was designed and trained using leave-one-patient-out cross-validation [27]. The 3D convolutional kernels that comprised the modified 3D inception V1 modules were of sizes 1×1×1, 3×3×3, and 5×5×5, and in total, the CNN architecture contained 2 inception modules, 3 convolutional layers, and 2 fully connected layers using the 3D patch input size of 25×25×91 (height × width × spectral bands) for un-processed HSI patches and 25×25×61 for pre-processed HSI patches. The total number of HSI samples in the cross-validation group was 647,000 normal HSI pixels and 877,000 cancer HSI pixels, calculated from the binary masks excluding glare, from the 21 tissue specimens from 7 patients. Each

HSI pixel served as the center for an image patch (height × width × spectral bands). Drop-out was employed to avoid over-fitting in the CNN models.

In total, the CNN described was separately trained using 4 different scenarios using the information from two versions of the binary ground truth masks and the two proposed processing methods, which produces different models: first, CNN trained without specular glare patches (trained separately both with and without added pre-processing of spectral data), and second, CNN trained with specular glare patches (trained separately both with and without added pre-processing of spectral data). The 95% confidence intervals were calculated using a bootstrapping method by sampling 1000 pixels from each class with replacement from each patient and calculating the area under the curve (AUC) of the receiver-operator characteristic; the method was performed 1000 times for each patient and the 2.5 and 97.5 percentiles were reported. The saved, trained models from the patients from the previous work were used to classify the HS images of the 5 testing group patients that underwent serial histological sectioning for surgical margin variation with depth evaluation. The probabilities of all models were averaged per patient to obtain qualitative probability heat-maps, scaled from 0 to 1, where 0 represents normal class and 1 represents high probability that the tissue belongs to the cancer class.

3.2.7.3 HELICoiD Algorithms

The results of the CNN classification method and the generated probability maps were compared to the results obtained by a machine learning pipeline previously developed for intraoperative detection of brain cancer using HSI [15, 167, 193-195]. In summary, a spatial-spectral classification algorithm, here referred to as HELICoiD (Figure 3-5a), was

implemented using a classification map obtained by a SVM classifier that is spatially homogenized by employing a combination of a one-band-representation obtained from the first principal component analysis (PCA) decomposition through a K-nearest Neighbors (KNN) filtering method (Figure 3-5a, part A). After that, the result of the KNN filtering is merged with an unsupervised segmentation map generated by a hierarchical K-means (HKM) algorithm through a majority voting (MV) method [169]. The result of this algorithm is a classification map that includes both the spatial and spectral features of the HS images. In addition, for this application, the KNN filtering is applied again to the MV probabilities and the PCA one-band-representation to homogenize the results (Figure 3-5a, part B).

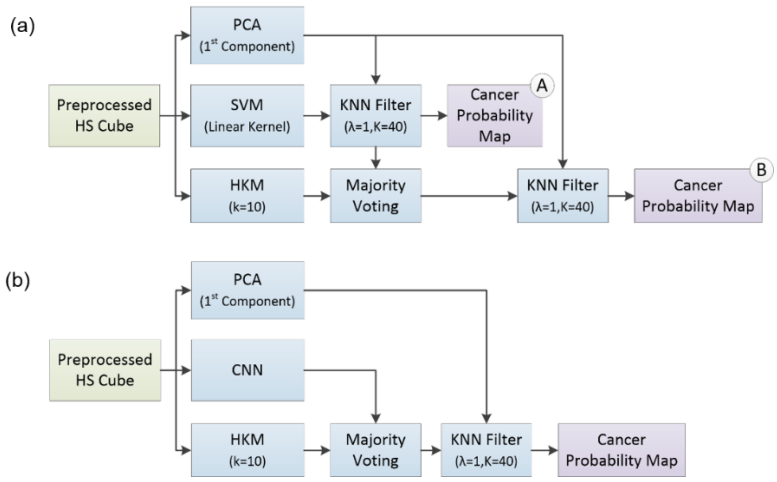

Figure 3-5. Block diagrams of the proposed classification frameworks. (a) HELICoiD algorithm with the additional KNN filter. (b) Pipeline of the mixed algorithm.

A component of the HELICoiD algorithm was isolated and referred to as spatial-SVM, which uses both the spectral and spatial components of HSI for machine learning through a combination of PCA+SVM+KNN. This spectral-spatial implementation of SVM is performed with and without the additional pre-processing pipeline to be used as a

surrogate for direct comparison to the CNN with and without pre-processing. Furthermore, a pipeline that combines the CNN architecture with the entire HELICoiD algorithm was proposed. In this case, the spatial-spectral stage of HELICoiD (PCA+SVM+KNN) is replaced by the CNN architecture trained with the pre-processed data (Figure 3-5b). Both HELICoiD and CNN+HELICoiD algorithms use the pre-processed HS data as input.

In summary, we present 6 algorithms for investigation of HSI machine learning: the CNN (with and without pre-processing), the spatial-SVM (with and without pre-processing), the HELICoiD algorithm, and finally the CNN+HELICoiD algorithm. These 6 machine learning algorithms were tested first on the group of experiments using the binary ground truth masks that exclude specular glare, and again the experiments were performed with the binary ground truth masks that include specular glare. The 6 algorithms were compared using box plot distributions of the values and median AUC with 95% confidence intervals, and, a paired, one-tailed t-test on the classification results was used to determine statistical differences with a 0.05 threshold for significance. The quantitative classification results of 7 patients are reported in Table 3-2, obtained by using leave-one-patient-out cross-validation. In addition, like the CNN method, the saved, trained models from these patients were used to classify 5 tissue specimens that were imaged with HSI and underwent serial histological sectioning for cancer margin variability evaluation. The probabilities of all models were averaged per patient to obtain qualitative probability heat-maps, scaled from 0 to 1, where 0 represents normal class and 1 represents high probability that the tissue belongs in the cancer class.

3.3 Results

3.3.1 Deformable Registration from Histology to HSI

Seventeen patient tissue samples were registered according to the described methods, divided into two groups: HNSCC (n=11) and thyroid (n=6). As a representative patient, Figure 3-6 shows a glossal SCC cancer margin with a needle-bored hole for TRE evaluation.

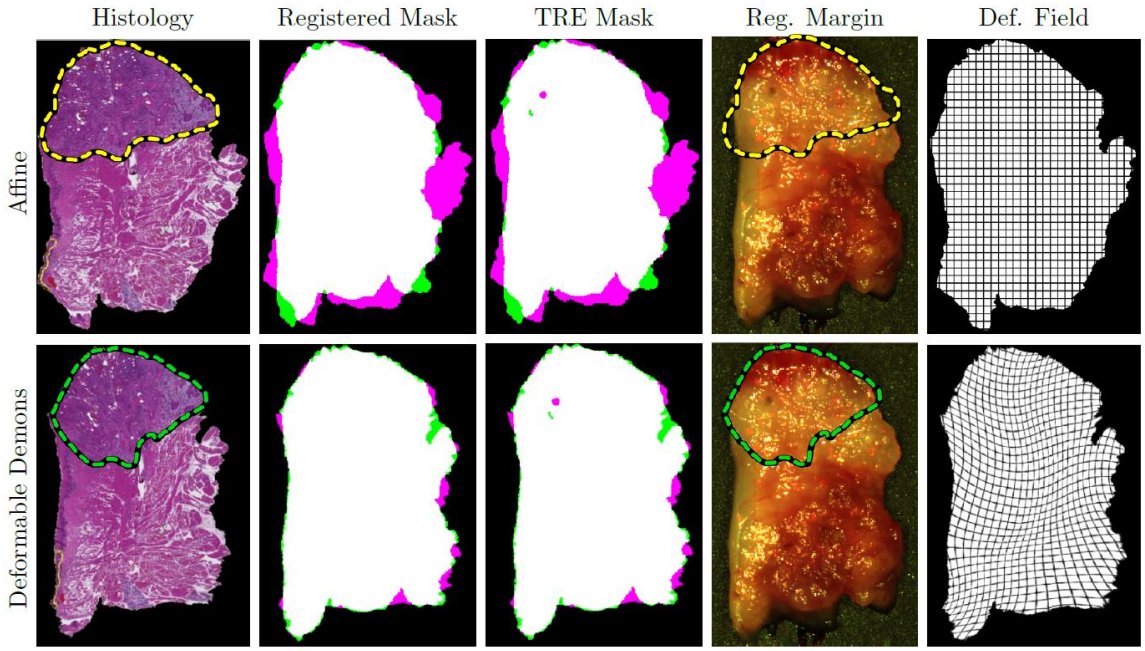

Figure 3-6. Results of affine and deformable demons registration of the cancer-normal margin on glossal SCC tissue sample with needle-bored holes. The cancer margin is outlined by a head & neck pathologist. The transformed margin from affine registration is outlined in yellow, and the margin from demons registration is outlined in green. The fixed HSI mask (green) and moving histology mask (magenta) overlap area is shown white. The TRE for this sample is evaluated using needle-bored holes shown on the masks in the center column. The deformation fields of the registration methods are shown on the right.

Affine registration alone of the representative tissue samples is insufficient to establish the cancer margin on the HSI, but the Demons-based method achieved successful alignment. Table 3-1 shows the complete experimental results. For the HNSCC group only,

Demons registration produced better matched images (TRE of 0.43 ± 0.16 mm) with more consistent results between tissue samples compared to affine registration alone (TRE of 0.51 ± 0.23 mm). Moreover, the result for HNSCC Demons registration was statistically significant (p=0.02) compared to affine, using a one-tailed, paired Student's t-test. In addition, for HNSCC, the proposed deformable pipeline had a better image overlap compared to affine alone, DSC of 0.98 versus 0.93, which was also statistically significant (p<0.001).

Table 3-1. Targeted registration error (TRE) and Dice similarity coefficient (DSC) for the registration between hyperspectral images and digitized histological images.

Group	Registration Method	TRE (mm)	DSC (%)
HNSCC	Affine	0.51 ± 0.23	93 ± 2
N=11	Deformable Demons	0.43 ± 0.16*	98 ± 2*
Thyroid	Affine	0.92 ± 0.45	91 ± 2
N=6	Deformable Demons	0.98 ± 0.42	98 ± 2*

*Statistically significant compared to affine registration (p<0.05)

The efficacy of both registration methods, affine and Demons, differed significantly for the thyroid group. On average the error for Demons registration of thyroid samples (TRE of 0.98 ± 0.42 mm) was comparable with affine registration alone (TRE of 0.92 ± 0.45 mm). However, the proposed deformable pipeline had a better image overlap for thyroid samples compared to affine alone, DSC of 0.98 versus 0.91, which was statistically significant (p<0.001). However, Demons registration did not show an advantage over affine registration for the thyroid samples.

Using a one-tailed, unpaired Student's t-test, the results of average error for HNSCC Demons registration were compared to thyroid sample Demons registration, and the result was statistically significant ($p<0.001$). This indicates that the registration of thyroid samples induces a substantially larger amount of error when compared to HNSCC samples, most likely due to less structural integrity of the glandular thyroid tissue.

3.3.2 Comparison of Machine Learning Methods with Specular Glare

Quantitative results from the leave-one-patient-out cross-validation, using both the ground-truth regions that include glare pixels and the sub-sampled masks that include only ideal quality regions that exclude glare, show that the CNN-based classifier group outperformed the SVM-based classifier group using the average area under the curve (AUC) of the receiver operator characteristic (ROC), as shown in Table 3-2 and Figure 3-7. The results are reported using seven-fold cross-validation to validate on all 7 patients.

Table 3-2. Results of inter-patient cross-validation of SCC versus normal, obtained using the leave-one-patient-out method. Average AUCs reported with bootstrapped 95% confidence interval.

Ground Truth	Classifier	Average AUC [95% CI]
Excluding Glare	spatial-SVM	0.71 [0.68, 0.74]
	CNN	0.86 [0.82, 0.89]
	spatial-SVM (Pre-processed HSI)	0.82 [0.80, 0.84]
	CNN (Pre-processed HSI)	0.84 [0.81, 0.86]
	HELICoiD (Pre-processed HSI)	0.82 [0.79, 0.84]
	CNN+HELICoiD (Pre-processed HSI)	0.82 [0.79, 0.85]
Including Glare	spatial-SVM	0.69 [0.67, 0.71]
	CNN	0.73 [0.71, 0.76]
	spatial-SVM (Pre-processed HSI)	0.76 [0.74, 0.77]
	CNN (Pre-processed HSI)	0.78 [0.76, 0.81]
	HELICoiD (Pre-processed HSI)	0.79 [0.77, 0.81]
	CNN+HELICoiD (Pre-processed HSI)	0.81 [0.80, 0.83]

When classification is performed of only ideal quality pixels (obtained from the sub-sampled mask; see Figure 3-1), the results indicate that additional pre-processing of the spectral signature and addition of the HELICoiD method and KNN filtering do not significantly improve the results compared to only using the CNN, see Figure 3-7a. The average AUCs for the CNN groups are 0.86, 0.84, and 0.82 for the CNN, CNN with pre-processed input data, and CNN+HELICoiD method, respectively. The average AUCs for the SVM-based groups are 0.71, 0.82, and 0.82 for the spatial-SVM without and with preprocessed input data, and HELICoiD method, respectively. The 95% confidence intervals overlap for all groups except for basis spatial SVM, as shown in Table 3-2. Additionally, all methods have a similar interquartile range and median distribution, see Figure 3-7.

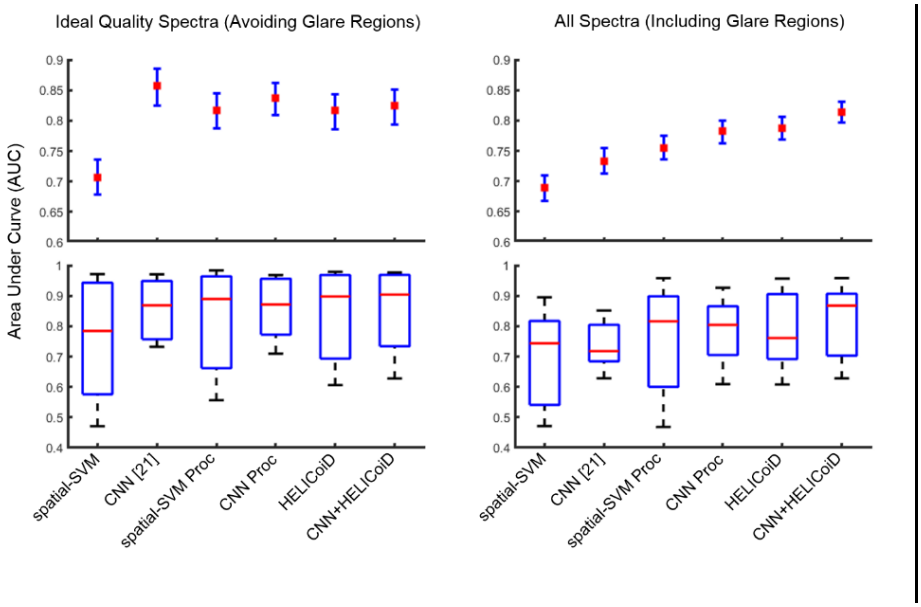

Figure 3-7. Results of inter-patient cross-validation of SCC versus normal, obtained using the leave-one-patient-out method. Top: average AUCs reported with 95% confidence interval. Bottom: box plots with the range in black, 75th and 25th percentile in blue, and median in red.

However, when the classification is performed over the entire HS image tissue area, which includes classification of specular glare pixels that contain more noise and variability, the CNN+HELICoiD method outperforms other methods tested with an average AUC of 0.81 for classification. This classification scheme including specular glare pixels represents a more realistic application of HSI. The CNN algorithm alone with or without pre-processing has an average AUC of 0.78 and 0.73, respectively, both which constitute a statistically significant decrease in performance compared the CNN+HELICoiD method ($p = 0.03$ and 0.04, respectively). The average AUC values range from 0.69 to 0.79 for the SVM-based groups, and the HELICoiD algorithm significantly outperforms the spatial SVM ($p = 0.01$). The additional pre-processing pipeline, described in Figure 3-4, only offers a statistically significant increase in performance for the spatial-SVM algorithm ($p = 0.02$), not the CNN ($p = 0.1$); see Table 3-2 for complete results with 95% confidence intervals. In summary, from cross-validation experiments, the best classification method was using the CNN as the input for the HELICoiD+KNN filtering method, and for comparison, using the spectral-spatial SVM+PCA+KNN for input to the HELICoiD algorithm instead of the CNN component yielded slightly lower results that were not statistically significant.

3.3.3 Variance of the Tumor Margin with Depth

For generalization and application, HS images from five tissue specimens from five patients with SCC comprised the hold-out testing group and were classified using the saved models that were trained and cross-validated using the 7 patient cross-validation group. The testing patients were classified by all cross-validation models and averaged to obtain qualitative probability heat-maps. Qualitative investigation of the five, testing patients,

classified with the CNN trained with pre-processed data alone, and the CNN+HELICoiD+KNN filtering method is shown in Figure 3-8. As shown in Figure 3-8, the CNN+HELICoiD+KNN technique performs better on the 5 testing patients, in agreement with the quantitative metrics from the cross-validation group. The SCC probability heat maps are shown with binary masks depicting the uncertainty and variation in the cancer margin with depth. Depending on the tissue, as demonstrated, the margin changes by about 1 mm depending on sectioning depth. Therefore, our qualitative results can be interpreted within the range of uncertainty of the ground-truth to provide more insight to the classification potential of machine learning methods using HSI for cancer detection.

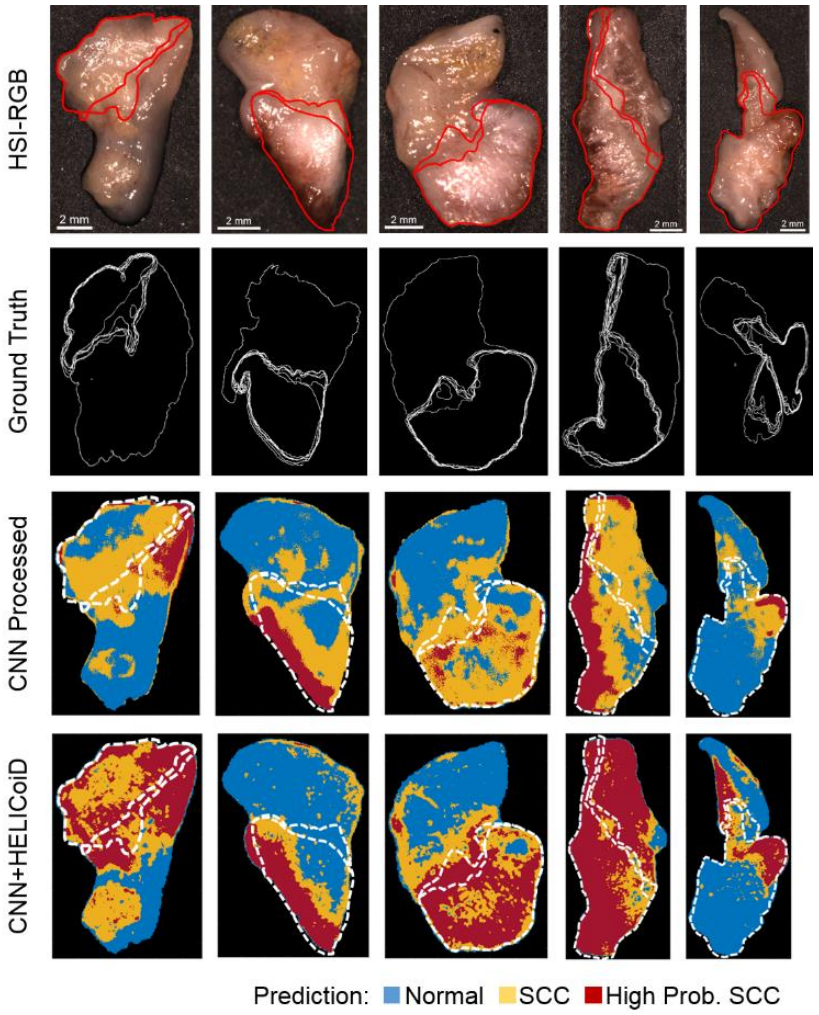

Figure 3-8. Representative results of binary cancer classification of the 5 testing patients. From left to right: HSI-RGB composite; histological ground truth showing variation in cancer margins with cancer area outlined; heat maps for cancer probability for CNN-HELICoiD+KNN and HELICoiD+KNN techniques. The extremes in the superficial cancer margin are overlaid on to the heat maps.

3.4 Discussion

In this work, we presented and quantified the combination of two state-of-the-art machine learning-based classification methods for HSI of *ex-vivo* head and neck SCC surgical specimens. In summary, there are two methods for generating SCC prediction probability maps: the first uses a CNN, and the second uses a combination of

SVM+PCA+KNN. After generation of the predicted cancer probabilities on a pixel level, the probability map is combined with a HKM unsupervised segmentation layer through a majority voting algorithm that determines the class belonging of the region of interest. Therefore, two distinct methods are compared using leave-one-patient cross-validation to obtain quantitative evaluation metrics. Additionally, the methods are tested on a group of HSI obtained from 5 SCC patients that underwent serial histological sectioning to evaluate the variation in the cancer margin with penetration depth of the light wavelengths.

The quantitative results of this paper suggest that when working with ideal quality pixels, such as the spectral signatures generated from flat surface tissue surfaces with no glare, the CNN techniques and spatial-spectral machine learning algorithms will perform with no significant difference and that no additional pre-processing is necessary. The average AUCs for these methods using the pre-processed input data range from 0.82 to 0.86 with overlapping confidence intervals. However, when the pixels classified contain noise, for example due to sloping of the tissue edges or specular glare from completely reflected incident light, additional spectral smoothing and additional HELICoiD+KNN filtering of the classifier improve classification results of the CNN. The best performing method tested was CNN+HELICoiD+KNN with an average AUC of 0.81 [0.80, 0.83]. These tested methods outperform the traditional spectral-spatial machine learning methods employed in this study. Therefore, the HELICoiD+KNN techniques using both the CNN and SVM+PCA+KNN for cancer probability maps performed best for the 7 cross-validation patients, so both were employed on the 5 patients testing set. One major limitation of the approach used in this paper was the small sample-size, and therefore, the proposed ML models could be prone to overfitting and lack generalization to larger testing

datasets. To investigate potential overfitting in this experiment, we analyzed the training and cross-validation accuracies for the CNN trained with pre-processed HSI data. The average training accuracy was 85%, and the average cross-validation accuracy with confidence interval was 79% [78, 81] for excluding glare and 74% [73, 76] for including glare. This result indicates that our models did not suffer substantially from overfitting.

Excluding glare, the CNN alone without pre-processing performs best. It is hypothesized that CNN and deep learning methods should be expected to outperform traditional ML techniques because a substantially large dataset should allow learning of variance, tolerance of noise, and remove the need for pre-processing. Also, the CNN+HELICoiD method performs only slightly better than the original HELICoiD; this could be produced because the CNN probability maps are out-weighted by the PCA and KNN filtering that are applied afterwards. After the inclusion of specular glare, the results change relative to each method, with the CNN with pre-processing out-performing the CNN without pre-processing. This outcome could be the result of the classification problem becoming more challenging compared to excluding glare, so the dataset is now too small for the CNN to perform well without pre-processing. Expanding the training dataset with more patients with specular glare and large amounts of noise may allow original CNN methods to outperform other techniques.

In order to test the general application of the proposed methods, HSI from 5 testing group patients with SCC were classified using the models that were trained and cross-validated using the 7 patient group. To qualitatively investigate these results, histological images of the 5 *ex-vivo* tissue specimens were obtained down to about 0.3 mm to determine how the superficial cancer margin may change with depth. As shown in Figure 3-8, the

CNN+HELICoiD+KNN technique performs best on the 5 patient SCC testing group, which was the same result obtained from the 7-fold quantitative cross-validation experiment.

The 5 patients' tumor-involved margin specimens were classified with the CNN trained with the fully pre-processed image patches that were extracted from the complete binary ground-truth mask including specular glare pixels. These probability maps are shown in Figure 3-8, and the CNN map is also used as the input for the HELICoiD+KNN filtering technique. In the top three rows of Figure 3-8, there are regions of glare with tissue specimens that are classified incorrectly as normal with the CNN, but the CNN+HELICoiD correctly classifies these regions of interest as SCC. It is also possible to observe that the CNN+HELICoiD method tends to over-predict SCC at regions of normal tissue near the cancer margin. We hypothesize that the CNN+HELICoiD technique outperforms the patch-based CNN alone because it can incorporate more local and regional spatial and spectral information to overcome the degrading effect of specular glare. Therefore, future work could involve the application of a fully-convolutional CNN for SCC detection on tissues with specular glare. This algorithm requires more data necessary for training as it produces labels that are end-to-end, a full pixel classification map for a full HSI input, so it would require the entire HSI dataset acquired for this project.

An additional aim of this investigation was to determine the variation of the superficial cancer margin through the depth of the tissue. From the 5 testing patient tissue specimens, it can also be seen that the superficial cancer margin of the *ex-vivo* tissue specimens can vary from the extreme near and far margins in range of 1 to 2 mm. These results allow interpretation of the cancer prediction probability maps with observed

variation in the ground truth. However, additional possible uncertainty may exist in the histological ground-truth. For example, if the angle of the sectioning plane is skewed from the tissue plane, then the ground-truth could be warped, which could lead to errors that cannot be corrected by deformable registration. In a previous work, we explored a pipeline of affine and deformable demons-based registration for alignment of the histological ground-truth to gross-level images of specimens, and it was determined via needle-bored control points that target registration error was 0.4 ± 0.2 mm [184]. The combination of these two main sources of uncertainty in the histological ground-truth for the *ex-vivo* tissue specimens allows for error propagation of up to 2.5 mm in the binary mask. This would greatly affect the results of pixel-wise evaluation metrics such as accuracy, sensitivity, and specificity.

Future work includes development and implementation of a new performance evaluation method to handle this margin uncertainty, for instance evaluating primary tumor clearance at several millimeter increments from the ideal tumor margin. Currently, the standard for surgeons' opinions on margin adequacy is 5 mm for head and neck SCC, with margins between 1 to 5 mm defined as 'close margins' [196]. However, some studies suggest that outcomes may be similar for margin of 2.2 to 5 mm [197]. It is evident that the question of margin adequacy is still being determined, and surgeons would be interested in performance of optical imaging methods at different margin distances. Therefore, the proposed, new error metric should extend to about 2 to 3 mm. Figure 3-9 demonstrates how the cancer-margin could be systematically eroded and used to determine accuracies at several distances from the cancer margin to better provide surgeons and physicians with a method for interpreting results for HSI studies, given the conclusions of this paper on the

uncertainty of the cancer-margin. The outcomes of this work suggest the use of this new error metric for future studies.

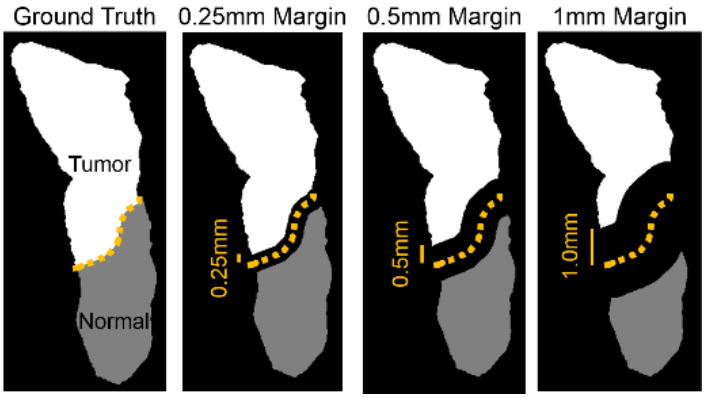

Figure 3-9. Proposed evaluation metric demonstrates millimetric, systematic cancer-margin erosion that could provide performance updates at several distances from the cancer margin to provide physicians with an easily interpretable method for understanding results of HSI studies.

Another avenue of future work involves the rethinking of the definition of normal tissues in the tissue specimens of the tumor-involved margin. In oral SCC, it has been studied that normal tissue directly adjacent to the primary SCC is molecularly distinct from normal tissues farther from the SCC, and additional resection to this extended margin may lead to increased disease-free survival and reduced local recurrence [177, 178]. Moreover, this result was obtained by investigation that NBI at 400–430 nm and 525–555 nm reveals changes in normal tissue that correlate to significantly different levels of micro-RNA epigenetic regulation compared to primary tumor and normal tissue that are not visible under white-light alone [175, 177, 178]. Therefore, it may be possible to extract micro-RNA expression levels to determine a molecular ground-truth for certain tissues employed in this study, as Liu *et al.* have demonstrated that micro-RNAs can be reliably obtained from formalin-fixed paraffin embedded tissue samples [198].

3.5 Conclusion

This chapter investigated the effects of specular glare, noise, blurring, and tissue-edge sloping artifacts on hyperspectral imaging-based cancer detection. It explored the potential of hyperspectral imaging and machine learning for the detection of head and neck cancer in a limited subset of tissue specimens. According to the experimental results, the CNN seems to be more robust against the environmental conditions of the acquired images and provides better classification results even without data pre-processing. Additionally, another objective was to evaluate the general efficacy on example test cases with uncertainty in the ground truth as the superficial cancer margin varies with penetration depth. This was tested by serially sectioning the tissue samples in the testing group to reveal the variation of the cancer margin through the depth of the tissue. This determined that the validity of the top section alone as the ground truth may be limited to 1-2 mm, suggesting an alternative approach for obtaining performance metrics should be developed. All the above factors were necessary to explore and understand the potential of HSI in the operating room. This chapter lays the foundation for the systematic limitations of the proposed experimental design that will be detailed in the following chapters, which will detail deep learning methods employed in sufficiently large patient datasets for training, validation, and testing to understand the potential role of HSI for clinical translation.

CHAPTER 4. HEAD & NECK SCC DETECTION IN 102 PATIENTS WITH HSI

This chapter details the work of using HSI for SCC detection on all SCC tissues collected in our dataset. Surgical resection of head and neck (H&N) squamous cell carcinoma (SCC) may yield inadequate surgical cancer margins in 10 to 20% of cases. This chapter investigates the performance of label-free, reflectance-based hyperspectral imaging (HSI) and autofluorescence imaging for SCC detection at the cancer margin in excised tissue specimens from 102 patients and uses fluorescent dyes for comparison. Fresh surgical specimens (n = 293) were collected during HNSCC resections (n = 102 patients). The tissue specimens were imaged with reflectance-based HSI and autofluorescence imaging and afterwards with two fluorescent dyes for comparison. Deep learning tools were developed to detect SCC with new patient samples (inter-patient) and machine learning for intra-patient tissue samples. The performance was estimated in mm increments circumferentially from the tumor-normal margin. In intra-patient experiments, HSI classified conventional SCC with an AUC of 0.82 up to 3 mm from the cancer margin, which was more accurate than proflavin dye and autofluorescence (both $p < 0.05$). Intra-patient autofluorescence imaging detected human papilloma virus positive (HPV+) SCC with an AUC of 0.99 at 3 mm and greater accuracy than proflavin dye ($p < 0.05$). The inter-patient results showed that reflectance-based HSI and autofluorescence imaging outperformed proflavin dye and standard RGB images ($p < 0.05$). In new patients, HSI detected conventional SCC in the larynx, oropharynx, and nasal cavity with 0.85–0.95 AUC score, and autofluorescence imaging detected HPV+ SCC in tonsillar tissue with 0.91

AUC score. This work demonstrates that label-free, reflectance-based HSI and autofluorescence imaging methods can accurately detect the cancer margin in ex-vivo specimens within minutes. This non-ionizing optical imaging modality could aid surgeons and reduce inadequate surgical margins during SCC resections.

4.1 Introduction

Surgery is often the primary treatment for head and neck squamous cell carcinoma (HNSCC) [199]. Primary surgery is the modality of choice for resectable oral cavity cancers and late stage disease of the larynx and hypopharynx [200]. Management of locally advanced SCC may also require a multimodal approach with adjuvant chemoradiation therapy [199, 201]. Nearly 90% of cancers of the upper aerodigestive track of the head and neck are SCC [2, 139]. Depending on the extent of the disease, radiation therapy or chemotherapy alone may be the primary curative modality selected, such can be the case with unresectable, recurrent, or metastatic cancers and also with cases known to be susceptible to chemoradiation [199, 200]. Human papilloma virus (HPV) is an identified cause of SCC, and the most common location for HPV positive (HPV+) SCC is the oropharynx, with nearly 60% of oropharyngeal SCC cases being HPV+ [199, 202]. Approximately two-thirds of patients with HNSCC present with stage III or IV advanced disease [141]. Adequate surgical removal of the primary SCC is vital to successful patient outcomes, improved quality of life, survival, and reduced recurrence [140, 196]. Surgeons can use pre-operative imaging, such as CT or MRI, for planning, but during the surgery, surgeons rely on experience, visual cues, and tactile palpation to determine the extent of the disease. Excised samples and tissue biopsies can be sent for pathological analysis and consultation to determine if the cancer has been sufficiently resected [196, 203, 204].

Intraoperative pathologist consultations (IPCs) can be time-consuming and may not fully reflect the extent of the disease due to limitations in tissue sampling and preparation. While the overall accuracy of frozen-sections in IPCs is upwards of 97%, the accuracy for challenging cases, such as positive and close margins, ranges from 71 to 92%, with sensitivities reported as low as 34 to 77% [203-207]. These errors can compound, leading to reported positive margins in up to 12% and close margins in up to 19% of HNSCC surgeries, despite having negative frozen sections during IPC [203-207]. For example, in oral cavity SCC, up to 30% of patients have positive margins after surgery [201, 208].

The task of surgical guidance for SCC resections in the head and neck has been explored with increasing volume in the past five years using several imaging modalities coupled with machine learning [209]. Some methods propose using fluorescently-tagged monoclonal antibodies that require intravenous administration but have specific optical signatures in the near-infrared (NIR) spectrum, with successful outcomes of studies with 21 patients [201, 208], and other methods utilize topical fluorescent dyes for targeting SCC [33, 210]. Label-free optical imaging methods that utilize only narrow bands in the blue and green visible spectrum have also demonstrated success at delineating oral SCC margins in-vivo in studies with 20 patients [177, 178].

Hyperspectral imaging (HSI) is an emerging technology in biomedicine [14] and has been used for cancer detection studies both ex-vivo and in-vivo [211, 212]. HSI has been utilized for brain cancer detection in-vivo using machine learning algorithms and an optimized, clinical workflow for neurosurgeons [166, 168]. Additionally, HSI has been proposed for laparoscopic cancer detection in colorectal surgeries with demonstrated potential [22, 129].

Our group reported proof-of-concept studies on HSI for the detection of head and neck SCC in fresh surgical specimens from human patients [33, 147]. In our previous pilot studies with HSI, manually selected regions of interest (ROIs) were classified, and image preprocessing was used to remove specular glare pixels before tissue classification [33, 210]. In our other works [113, 172, 179], deep learning algorithms were developed for HSI tissue classification in both whole primary-tumor specimens and at the SCC cancer-normal margin, but only in limited sample sizes from 21 to 29 patients employing cross-validation. Previous works from other groups focus on SCC detection in excised tongue SCC specimens, using both proof-of-concept ROI-based detection of SCC in 7 specimens [180] and tumor semantic segmentation of the entire cancer-margin specimens [148] with promising results in leave-one-patient-out cross-validation experiments of 14 patients.

In this large study of 293 tissue specimens from 102 patients with SCC, we develop deep learning methods to classify the whole tissue specimens instead of ROIs and thus further investigate the full potential of label-free HSI-based imaging methods for SCC detection. This is the first work to conduct fully-independent training, validation, and testing directly of the SCC tumor margin with a large patient dataset (N = 102 patients), divided into conventional, keratinizing SCC with variants (N = 88) and HPV+ (N = 14) SCC cohorts. The tissues represent a variety of anatomical sites to give an accurate assessment of the feasibility of label-free, non-contact, and non-ionizing HSI-based imaging modalities for SCC detection. Additionally, this is the first study to investigate and quantify HSI-based methods for HPV+ SCC detection directly. It is hypothesized that deep learning algorithms can be developed to enable label-free HSI-based methods, namely reflectance-based HSI and autofluorescence imaging, to perform with substantial accuracy

to provide meaningful information to guide complete surgical resections. Furthermore, it is hypothesized that label-free HSI-based methods will outperform the fluorescent dye-based methods due to lack of target specificity with sufficient signal-to-noise in SCC tissues. The results of this study will inform if HSI and other fluorescence imaging modalities can be expected to provide specific benefits to cancer margin detection during SCC resection surgeries.

4.2 Materials

4.2.1 Head and Neck SCC Patient Dataset

Patients with head and neck SCC undergoing routine surgery at the Emory University Hospital Midtown (EUHM) were recruited by providing informed, written consent to the research coordinator, who de-identified the patient data. All methods and procedures were approved by the Emory University Institutional Review Board (IRB) under the Head and Neck Satellite Tissue Bank protocol. Fresh, ex-vivo surgical specimens were collected from the surgical pathology laboratory, making sure not to impede routine clinical service. Three tissue samples from each patients' gross tissue specimen were collected: a tissue specimen of the primary tumor (T), an all normal tissue (N), and a specimen at the tumor-involved cancer margin (TN). The specimens were transported to an imaging laboratory to perform gross-level optical imaging of the ex-vivo specimens for SCC detection [33, 113, 210]. The median sizes (height × width) of the tissues were 9×6 mm, 10×7 mm, and 9×5 mm for the T, TN, and N tissues, respectively, with an approximate tissue depth of 2 to 3 mm. In total, 293 tissue specimens were collected from 102 patients with head and neck SCC to be included in this study.

A pathologist with expertise in head and neck cancer categorized the tissue samples collected for this study into two groups per cancer subtype: Conventional SCC with variants and HPV+ SCC. The conventional SCC group (N = 88 patients) was comprised of conventional, keratinizing SCC (N = 85), adenosquamous carcinoma (N = 1), basaloid SCC (N = 1), and spindle cell SCC (N = 1). The HPV+ SCC group consisted of 14 patients that were all identified as p16 marker positive using immunohistochemistry, and one was a neuroendocrine SCC. Table 4-1 shows the breakdown of tissue samples for different locations of primary tumors along with other patient characteristics and cancer properties. For this study, we defined the oral cavity as broadly consisting of non-tongue structures inside the oral cavity: Oral mucosa (N = 3), gingiva (N = 5), floor of mouth (N = 12), retromolar trigone (N = 4), maxillary (N = 2), and mandibular surfaces (N = 9). Although it is not the clinical standard, tongue specimens were presented separately because it was the single largest anatomical site in this study, and the excised specimens we acquired from the tongue were typically very distinct, both optically and anatomically, with thick surface epithelium and large amounts of skeletal muscle, compared with the rest of the oral cavity cases. The pharynx consisted of cases originating in the oropharynx (N = 12 total; 10 HPV+ and 2 HPV-), hypopharynx (N = 4), and N = 3 cases of HPV+ tonsillar SCC. One patient's primary location was unknown, only nodal HPV+ neck mass was available.

Table 4-1. Demographics and cancer properties for the patients recruited for this study. Values are reported for the two cohorts, conventional SCC with variants and p16+ SCC, separately and combined. Percentages for conventional are out of 88 patients, HPV+ out of 14 patients, and combined out of 102 patients. TNM staging was not available for one patient in the HPV+ cohort. All cases were M0. Tobacco history represents current or past smoking or chewing tobacco history.

Property	Conventional SCC (*N* = 88)		HPV+ SCC (*N* = 14)		All SCC (*N* = 102)	
	No.	%	No.	%	No.	%
Demographics						
Mean Age (y.o.)	64.5	-	58.1	-	63.6	-
Male	59	67%	12	86%	71	70%
Female	29	33%	2	14%	31	30%
Tobacco History	58	66%	5	36%	63	62%
Primary Location						
Oral Cavity	35	40%	0	0%	35	34%
Tongue	19	22%	0	0%	19	19%
Oropharynx	2	2%	13	93%	15	15%
Hypopharynx	4	5%	0	0%	4	4%
Larynx	19	22%	0	0%	19	19%
Nasal Cavity	4	5%	0	0%	4	4%
Maxillary Sinus	5	6%	0	0%	5	5%
Unknown	0	0%	1	7%	1	1%
Cancer Stage						
pT1	3	3%	2	14%	5	5%
pT2	7	8%	6	43%	13	13%
pT3	16	18%	2	14%	18	18%
pT4	62	70%	3	21%	65	64%
Avg. T Size (cm)	4.4	-	3	-	4.2	-
N+	53	60%	5	36%	58	57%
Histologic Grade						
G1	8	10%	-	-	8	8%
G2	60	71%	-	-	60	59%
G3	16	19%	-	-	16	16%
IPC (Averages)						
IPC/Surgery	2.1	-	2.0	-	2.1	-
Time/IPC (min)	41	-	42	-	41	-
Tissues/Surgery	3.5	-	2.6	-	3.4	-
Time/Tissue (min)	25	-	33	-	25	-

4.2.2 Hyperspectral Imaging

The HSI were acquired of the gross-level surgical specimens using a Maestro spectral imaging system (Perkin Elmer Inc., Waltham, Massachusetts), which captured 2D images at full spatial resolution using a 16-bit charge coupled device and stepped through the spectral bands using a liquid crystal tunable filter [33, 113, 210]. The spatial resolution of the HS sensor was 1040 by 1392 pixels, which corresponds to a specimen-level resolution of 25 μm per pixel. The HSI were captured from 450 to 900 nm in 5 nm spectral bands to produce a hyperspectral data cube (hypercube) with 91 spectral bands. The average imaging time for acquiring a single HSI at this resolution was about one minute. The hyperspectral data were normalized in each spectral band individually by subtracting the inherent dark current (imaging with a closed camera shutter) and dividing by a white reference disk. An RGB composite image was generated from the normalized hypercube by applying a Gaussian kernel in each color region. Figure 4-1 shows a representative patient tissue specimen and average spectral signatures of SCC and normal tissues.

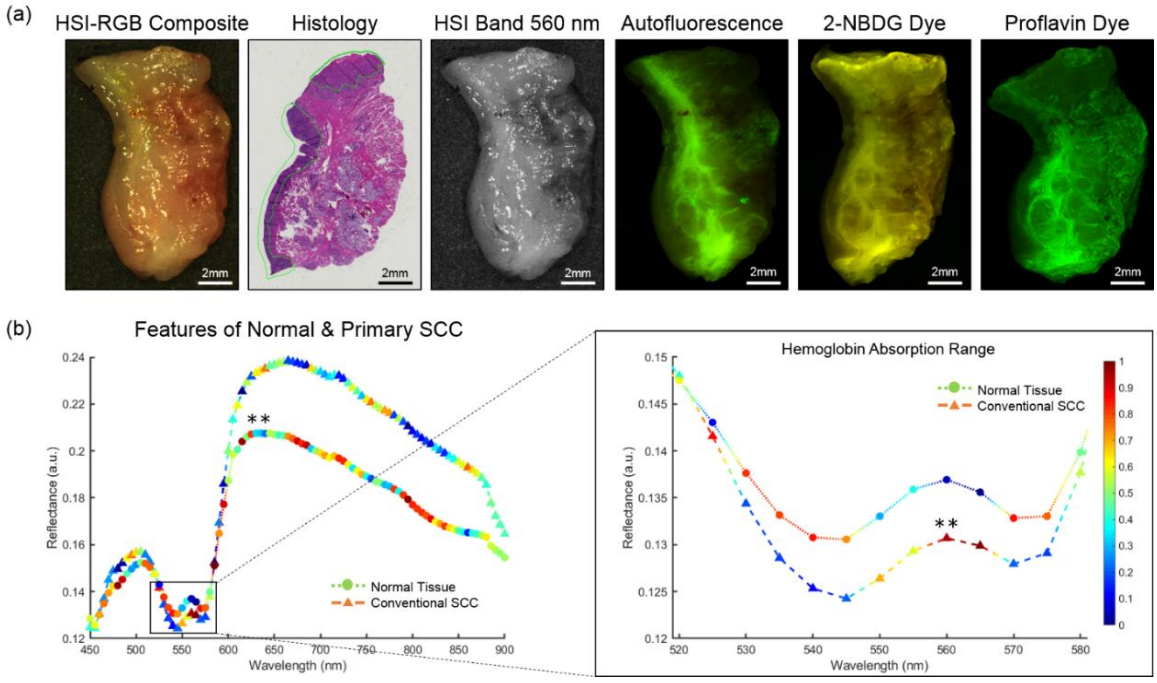

Figure 4-1. A representative cancer-involved tissue specimen of conventional, keratinizing SCC of the mandibular gingiva. (a) From left to right: RGB image made from HSI. The histological image, which serves as the ground truth, has SCC annotated in green. HSI single band at 550 nm. Fluorescence imaging modalities of the same specimen; (b) Spectral feature saliency from CNN gradients of correctly classified HSI for conventional SCC and normal upper aerodigestive tract tissues. Left: Full spectra from 450 to 900 nm of SCC and normal tissues. Symbol colors represents the relative, scaled importance of the spectral feature for making the correct prediction of cancer or normal from the HSI (0 is low saliency; 1 is high saliency). Right: Spectral cutout from 520 to 580 nm, corresponding to the hemoglobin range. The double asterisk (**) indicates that statistically significant differences ($p < 0.01$) were observed in reflectance values between SCC and normal for all spectral bands (450 to 900 nm). The most important spectral feature for correctly predicting SCC in HSI was the oxygenated hemoglobin peak at 560 and 565 nm.

4.2.3 Fluorescence Imaging

To compare the ability of HSI for cancer margin detection, several optical imaging modalities were acquired afterwards: A second label-free and two dye-based methods [33, 210]. Figure 4-1 shows a representative patient tissue specimen of all imaging modalities

and the histological ground truth. Autofluorescence imaging is a label-free imaging modality that captures the light emission from intrinsic fluorophores in biological tissue that are stimulated to fluoresce by external excitation. The autofluorescence images were produced by an excitation light source of 455 nm and a long-pass filter of 490 nm. The autofluorescence images were acquired from 500 to 720 nm in 10 nm increments to produce a hypercube of 23 spectral bands, and this imaging protocol was also used for the fluorescent dyes.

In addition to the two label-free modalities (HSI and autofluorescence), two dye-based fluorescence imaging modalities were acquired for SCC detection [33, 210]. A fluorescently tagged glucose molecule, 2-deoxy-2-[(7-nitro-2,1,3-benzoxadiazol-4-yl) amino]-D-glucose (2-NBDG), is a dye that produces a stronger signal measured from regions with higher metabolic glucose uptake, often associated with cancer regions. The tissue specimens were incubated for 20 minutes in a 160 μM 2-NBDG solution (Cayman Chemical, Ann Arbor, MI, USA) at 37 degrees Celsius, after which tissues were rinsed in 1× phosphate buffered solution (PBS) before imaging.

Proflavin dye images were acquired last before fixing the tissues [33, 210]. Proflavin is a DNA-binding fluorescent dye that has demonstrated utility for nuclear morphology visualization in a mouse model of oral carcinogenesis [213]. The effect of proflavin staining allows greater signal-to-noise in non-keratinizing tissues, as keratin is also a target of the dye [213, 214]. The tissue specimens were incubated for 120 seconds in a 0.01% proflavin solution (Sigma Aldrich, St. Louis, MO, USA) at room temperature, after which tissues were rinsed in PBS.

4.3 Methods

4.3.1 Histological Ground Truth and Registration

After acquiring the HSI, the fresh, ex-vivo tissue specimens were inked to preserve the optical imaging orientation, fixed in formalin, and paraffin embedded. Using a microtome, only the first, high-quality top section corresponding to the surface that was optically imaged was obtained, stained with hematoxylin and eosin, and digitized using whole-slide scanning at 40× objective [215]. The digital histology images from each specimen were annotated to outline the cancerous and normal areas by a board-certified pathologist with expertise in H&N cancer [33, 210].

The digital histology ground truth served as the gold standard for the optical imaging modalities. The histology ground truth image was registered in a semi-automated fashion according to a previously established pipeline of deformable registration to the gross-level HSI [184]. This registration was subject to errors in tissue-deformation, uncertainty in the cancer margin with depth, and off-plane slices that in total were estimated to be about 1 mm [184]. Moreover, the variation in photon penetration depth from the optical imaging modalities and the variation in the cancer margin throughout the depth of the tissue specimens was also estimated to create another 1 to 2 mm of error in the margin, according to our previous work [216]. Therefore, a systematic and objective method for calculating classification performance was implemented by removing the area near the cancer margin in millimeter increments and reporting all values [216]. The regions near the cancer margin are both included and excluded from performance calculations because the tissue near the margin can be degenerated or have undergone pre-cancerous

transformation. Here the registered cancer margin is referred to as the 'actual TN' margin, and mm increments estimated from the TN margin are identified. The 'actual TN' margin calculates performance metrics for all tissues right up to the pixels that comprise the interface of tumor and normal. For distance calculations for example, 'TN at 1 mm' represents that evaluation metrics are calculated from all distances up to 1 mm from the margin. The 'TN margin at 2 mm' is also reported, which calculates performance up to 2 mm from the margin.

4.3.2 Intra-Patient Experiments

Intra-patient experiments used a patient's known cancer and normal specimens to train a machine learning algorithm and simulated a personalized approach for SCC detection on-the-fly in the operating room. For intra-patient experiments, linear discriminant analysis (LDA) was used in ensemble to train, validate, and test the SCC data from the same patient. Each SCC patient with all tissue types (meaning a purely normal specimen, a specimen containing only primary tumor, and a specimen of the cancer margin) was included and divided into each cohort, conventional SCC ($N = 41$) and HPV+ SCC ($N = 6$). Despite collecting 102 patients for this study, only 47 fit this distribution of all three tissue types exactly. Independently, an ensemble LDA of 500 learners was trained and validated in 5 folds from each patient's tumor and normal samples. After each patient's model development, the patient's tumor-normal margin specimen was used as the testing data. The LDA method was selected because our previous work demonstrated that it outperformed other regression-based machine learning algorithms [33]. Training time for 5 cross-validated folds of one patient's model was about one to three minutes, depending on the size of regions selected for training, which is reasonable for simulating training of a

patient's data for HSI during surgery. All statistical analyses were performed using a paired, one-tailed t-test.

4.3.3 Inter-Patient Experiments

To explore the ability of HSI and fluorescence imaging modalities to detect SCC on patients fully-independent from algorithm development, two experiments were performed. The first experiment consisted of training the CNN on primary tumor (T) and all normal (N) tissues, while testing on T and N tissues from other patients. The second experiment consisted of training on primary tumor (T) and all normal (N) tissues, while testing only tumor-involved cancer margins (TN) tissues from other patients.

To perform these experiments, within each SCC cohort, patients were randomly divided into 5 folds, each fold served as the fully-independent testing group, while training and validation was performed on the patients in the remaining 4 folds, which allows test-level performance metrics for all patients in our dataset. For the conventional SCC cohort, each model from each fold was trained and validated on approximately 25,000 patches from 110 tissue specimens from 70 patients, and the independent testing group from each fold was approximately 50 tissues from 20 patients. This was performed once for each fold, until the entire cohort dataset, comprised of 70,000 patches from 255 tissue specimens from 88 patients was used as the testing group. For the HPV+ SCC cohort, training/validation was performed in the same fashion in 5 folds, until the entire cohort dataset of 16,000 patches from 38 tissue specimens from 14 patients was used as the testing group. All statistical analyses were performed using a paired, one-tailed t-test.

4.3.4 Convolutional Neural Network

For inter-patient experiments, a convolutional neural network (CNN) was developed to quickly and efficiently classify cancerous and normal tissues at the cancer margin. Due to the uniqueness of HSI data, the inception-v4 CNN architecture [217] was customized in several key ways to optimize the CNN to hypercube data in image-patches that were 25×25×*C*, where *C* is the number of spectral bands of each HS optical modality. The full CNN architecture schematic is presented in detail in Figure 4-2. The CNN was developed in TensorFlow on an Ubuntu machine running NVIDIA Titan-XP GPUs [192]. The early convolutional layers were modified to handle the selected patch-size and create smaller inception blocks that would allow for faster training and classification using the CNN. Training was performed up to 50 epochs; one epoch of training data ran for up to 1 hour using HSI; and deployment of the fully-trained CNN on a single GPU to classify a new HSI scene with hundreds of patches required only 25±10 seconds. The relative saliency of spectral features for correctly predicting SCC or normal in HSI, shown in Figure 4-1b and Figure 4-7, were extracted from the CNN using the class-activated gradients per the grad-CAM algorithm [218].

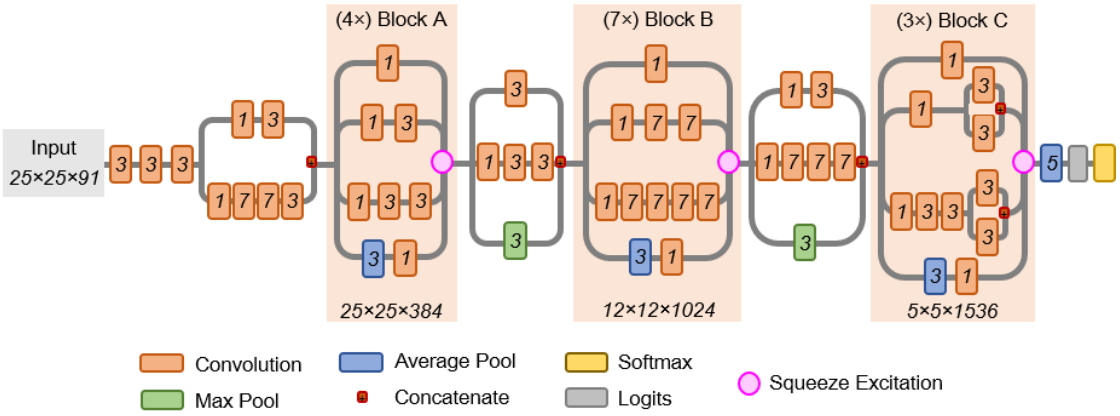

Figure 4-2. Schematic of the customized Inception V4 CNN architecture with added squeeze-and-excitation layers. The CNN was modified to operate on the 25×25×91 patch-size selected. The receptive field size and number of convolutional filters is shown at bottom of each inception block. The convolutional kernel size used for convolutions is shown in italics inside each convolution box.

4.3.5 Image Processing and Reconstruction

For all experiments, the 10 pixels, corresponding to 0.25 mm, at the edge of each tissue were discarded for performance calculations. Since the imaging protocol for tissue specimens required using a flat imaging surface, the tissue free edges created false curvature where the tissue was too thin to provide an adequate imaging signal. Implementation of the inter-patient CNN experiments involved a patch-based approach using a sliding window of size 25×25×91 and an overlap of 13 pixels. The overlapping regions of image-patches were averaged to produce a smoother result for calculation of the performance metrics of the inter-patient experiments.

4.3.6 Evaluation

To evaluate performance of the machine learning algorithms employed in the experiments for detecting SCC, the area under the curve (AUC) of the receiver operator

characteristic (ROC) curve was calculated. The AUC score was selected because it describes the accuracy at all possible thresholds of identifying the positive class and is not susceptible to errors when the classes are imbalanced. For each experiment, the optimal operating point on the ROC was calculated for the validation group data. This validation group threshold was used as the threshold for the testing group to best distinguish between cancer and normal, objectively. Using this threshold, the overall accuracy was calculated. Sensitivity, the ratio of true positives to total positive predictions, and specificity, the ratio of true negatives to total negative predictions, were also calculated and presented.

4.4 Results

4.4.1 Surgical Specimens Results

The accuracy of pathologist assistants in the surgical pathology department was calculated on their ability to identify the desired tissue specimen type (T, TN, or N) for research purposes. To obtain this performance, the label (prediction) given by the pathologist assistant (T, TN, or N) was compared to the ground-truth label from histology. This value is reported only to give an estimate of the difficulty of the task of positive margin and primary tumor specimen identification. It is important to note that these specimens were for research purposes only and do not attempt to reflect the accuracy of the clinical service in determining SCC during intraoperative guidance. The accuracy for identifying tissue specimens correctly was normal specimens with 92% accuracy, tumor-normal margin tissues with 95%, and primary tumor-only specimens with 60% accuracy. Figure 4-3a shows the breakdown by tissue specimen type. The most common reason for the misidentified tissue specimens was normal tissue in the primary tumor specimen. To

calculate the accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the specimen identification in surgical pathology, the TN tissues and predictions were separated into T and N components for calculation of true/false positives/negatives. For example, a true TN predicted as all T would count as both a true-positive and a false-positive; alternatively, a true T predicted as TN, would count as both a true-positive and a false-negative. The specificity and PPV were both 82%, the lowest metrics obtained, as the most common misidentification was tumor-margins as tumor-only, and overall an accuracy of 88% was determined for tissue type identification. Figure 4-3b details the complete performance metrics calculated and the number of each tissue components.

(a)

Surgical Path Prediction

Histology Ground Truth	T	TN	N	Total
T	**59**	2	0	61
TN	37	**94**	8	139
N	2	3	**88**	93
Total	98	99	96	293
Accuracy	60%	95%	92%	82%

(b)

Surgical Path Prediction

Histology Ground Truth	T	N	Total	
T	192	10	202	Sens.: 95%
N	42	193	235	Spec.: 82%
Total	234	203	437	
	PPV: 82%	NPV: 95%		Acc.: 88%

Figure 4-3. Accuracy of ex-vivo tissue samples acquired. (a) The types of tissue samples acquired were all normal (N), primary tumor (T) specimen, or specimen at the tumor-normal (TN) margin. Accuracies shown are for the desired tissue type in the column. (b) The performance metrics when TN tissues and predictions were separated into T and N components for calculation of accuracy, sensitivity, specificity, PPV, and NPV. For example, a true TN predicted as all T would count as both a true-positive and a false-positive; alternatively, a true T predicted as TN, would count as both a true-positive and a false-negative. Therefore, the 139 true TNs and 5 false predicted TNs are double counted, so the specimen total is now 437 instead of 293 specimens.

4.4.2 Intra-patient Results of Tumor-Involved Cancer Margins

The results of reflectance-based HSI intra-patient (N = 47) accuracy ranged from 80 to 90% for conventional SCC and 82 to 97% for HPV+ SCC. The results of reflectance-based HSI intra-patient AUC were 0.75 to 0.82 for conventional SCC and 0.77 to 0.91 for HPV+ SCC. Table 4-2 shows the complete results for HSI and the fluorescence imaging modalities. The accuracy for both SCC cohorts peaked at 3 mm from the cancer margin, but the number of patients decreased with increasing distance from the cancer margin because not all distances up to 3 mm could be estimated from all tissue specimens. For the conventional SCC cohort (N = 41), HSI outperformed autofluorescence, proflavin, and 2-NBDG at nearly all distances from the cancer-margin and significantly outperformed proflavin and autofluorescence ($p \leq 0.05$) close the margin (Figure 4-4a,b). The results for the HPV+ SCC cohort (N = 6) are not as conclusive as the conventional cohort because the limited number of tissue samples causes a discontinuous trend, with autofluorescence imaging being among the top performing modalities in terms of accuracy and AUC (Figure 4-4c,d), and the AUC obtained for autofluorescence imaging detection of HPV+ SCC was 0.99 at 3 mm, which was greater than other modalities ($p > 0.05$). In terms of accuracy, autofluorescence imaging significantly outperforms proflavin at 2.5 mm from the cancer margin ($p < 0.05$). Therefore, the label-free methods, HSI and autofluorescence, perform best for intra-patient testing at the cancer margin.

Table 4-2. Accuracy (Acc.) and area under the curve (AUC) results for all modalities performed using intra-patient experiments for conventional SCC and HPV+ p16-positive cohorts. Bolded values represent the greatest value in the column for each patient cohort.

Imaging Modality	SCC Cohort	1mm		2mm		3mm	
		AUC	Acc.	AUC	Acc.	AUC	Acc.
HSI	SCC, Conventional	**0.77**	**85%**	**0.80**	**88%**	0.78	**90%**
	SCC, HPV+	0.79	88%	0.69	88%	0.91	97%
Autofluorescence	SCC, Conventional	0.67	85%	0.72	85%	0.73	87%
	SCC, HPV+	0.81	89%	**0.84**	**95%**	**0.99**	**98%**
Proflavin	SCC, Conventional	0.73	82%	0.76	85%	0.73	89%
	SCC, HPV+	**0.85**	88%	0.82	94%	0.93	97%
2-NBDG	SCC, Conventional	0.76	84%	0.79	87%	**0.80**	89%
	SCC, HPV+	0.74	**90%**	0.69	95%	0.70	97%

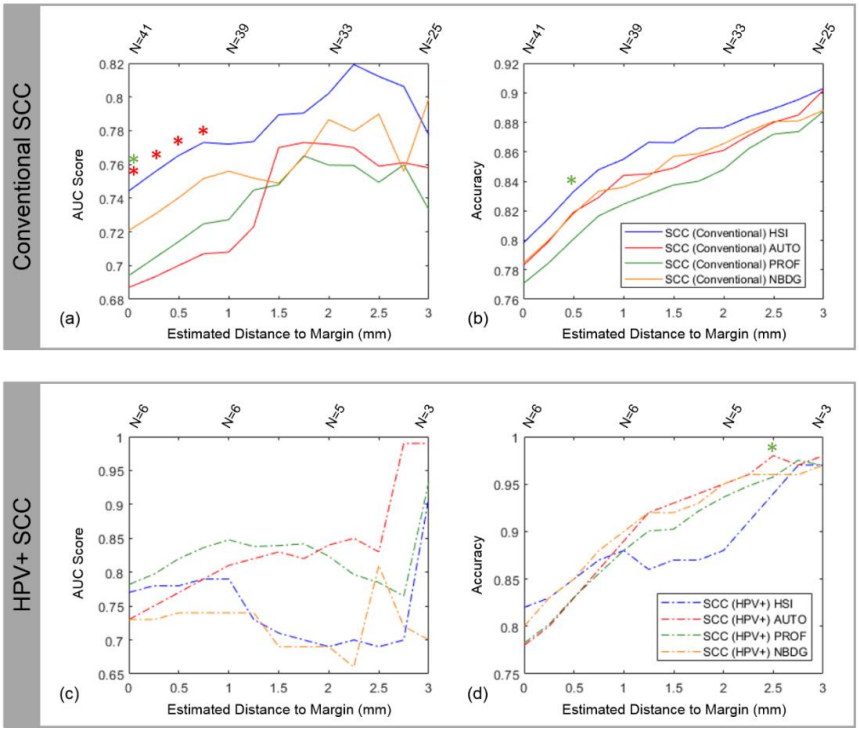

Figure 4-4. Results from intra-patient training and testing with LDA using HSI, autofluorescence, proflavin, and 2-NBDG. Results for Conventional SCC are shown in (a) AUC and (b) accuracy. Results for HPV+ SCC are shown in (c) AUC and (d) accuracy. Statistically significant results between imaging modalities are indicated by a colored asterisk. Sample size (N) is reported above plots to indicate that not all distances in mm can be estimated from each tissue specimen, so sample size decreases as the distance estimated increases, which causes noticeable jumps in the plots.

4.4.3 Inter-patient Results of Tumor vs Normal

For the conventional SCC cohort (N = 88), the results of the inter-patient experiments using only one-class specimens of the tumor-only (T) or normal-only (N) yielded similar median and average AUC scores for HSI and autofluorescence of 0.92 and 0.93 (median) and 0.87 and 0.86 (average). Both of HSI and autoluorescence outperformed the fluorescent dye-based techniques, 2-NBDG and proflavin, which had 0.88 and 0.87 (median) and both with 0.85 average AUC scores (all differences not significant with $p >$ 0.05). Figure 4-5c,d shows the median and average AUCs for all imaging modalities for the conventional SCC cohorts. For the HPV+ SCC cohort (N = 14), autofluorescence imaging yielded a median AUC of 0.86 and average AUC of 0.74, which was significantly more accurate than proflavin or RGB imaging (both $p < 0.05$). For tonsillar HPV+ SCC tissues (N = 3), the average SCC detection was 0.89 AUC score (Figure 4-6b,d). The grad-CAM algorithm [218] was used to show that the salient spectral features necessary for distinguishing SCC from normal across all anatomical sites encompassed both the visible and NIR spectrum (Figure 4-1b and Figure 4-7). In particular, it can be observed that normal tissues have more absorption and salient spectral features in the NIR range, indicating that normal tissues have greater fat, collagen, and water content than cancer [116]. The most important spectral feature for correctly predicting SCC with HSI was the oxygenated hemoglobin peak at 560 and 565 nm, which is correlated with increased metabolic activity. Additionally, the spectral signatures and spectral feature saliency maps are shown separated by anatomical location in Figure 4-7.

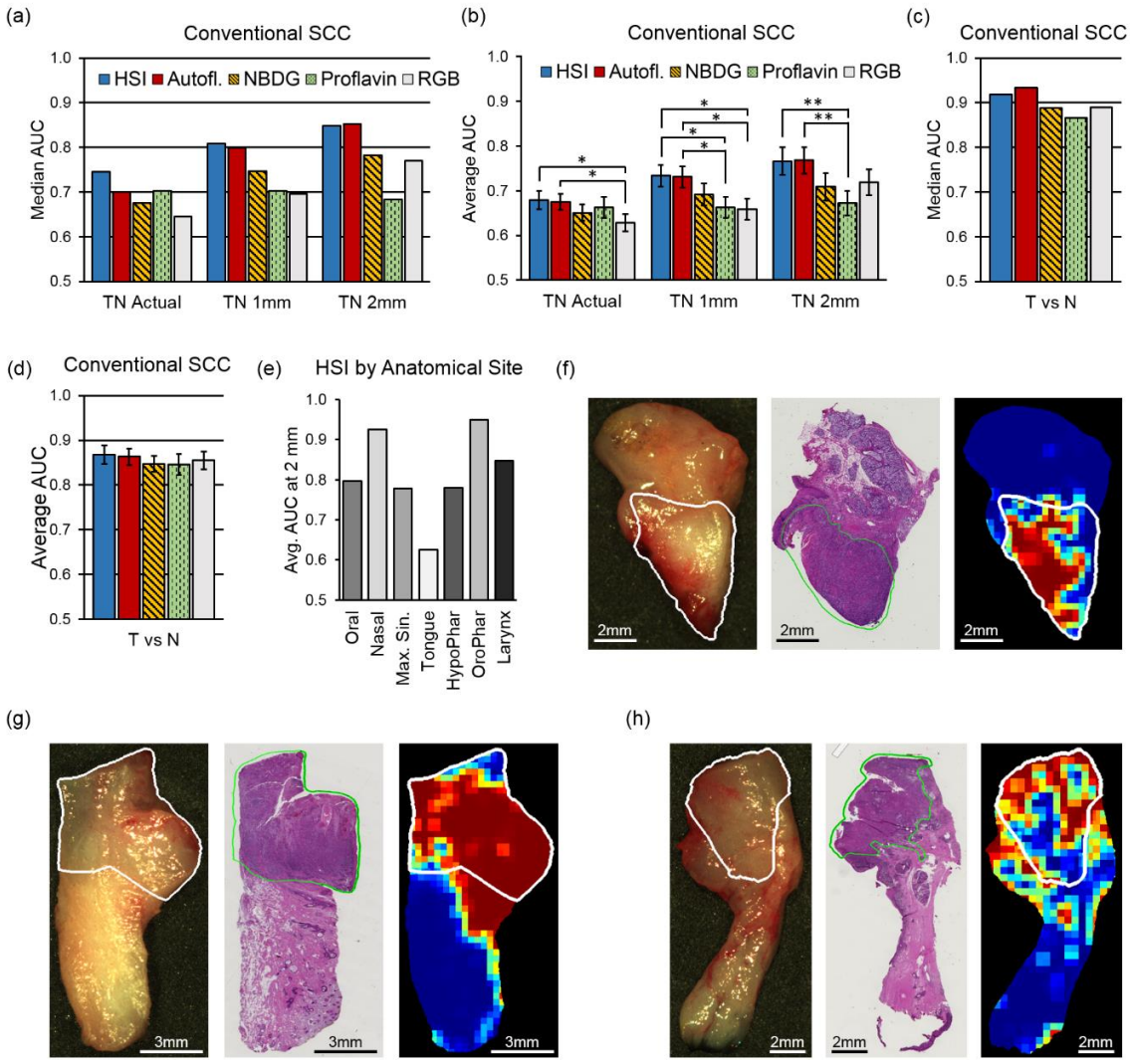

Figure 4-5. Median and average AUC results from inter-patient classification (a value of 0.5 corresponds to random guess). AUC values for the conventional SCC cohort: (a) Median AUC values for TN margin tissue specimens; (b) average AUC shown with SEM for TN margin specimens with statistical significance, shown as (*) for $p<0.05$ and (**) for $p<0.01$; (c) median AUC values for T and N whole tissue specimens; (d) average AUC shown with SEM for T and N whole specimens; (e) Average AUC at 2mm from the SCC margin using HSI across different anatomical sites; (f–h) representative patient examples of conventional SCC at the maxillary gingiva, nasal cavity, and larynx, respectively. From left to right: RGB made from HSI, histology ground truth, and predicted cancer heat-map. The white and green contours outline the SCC area.

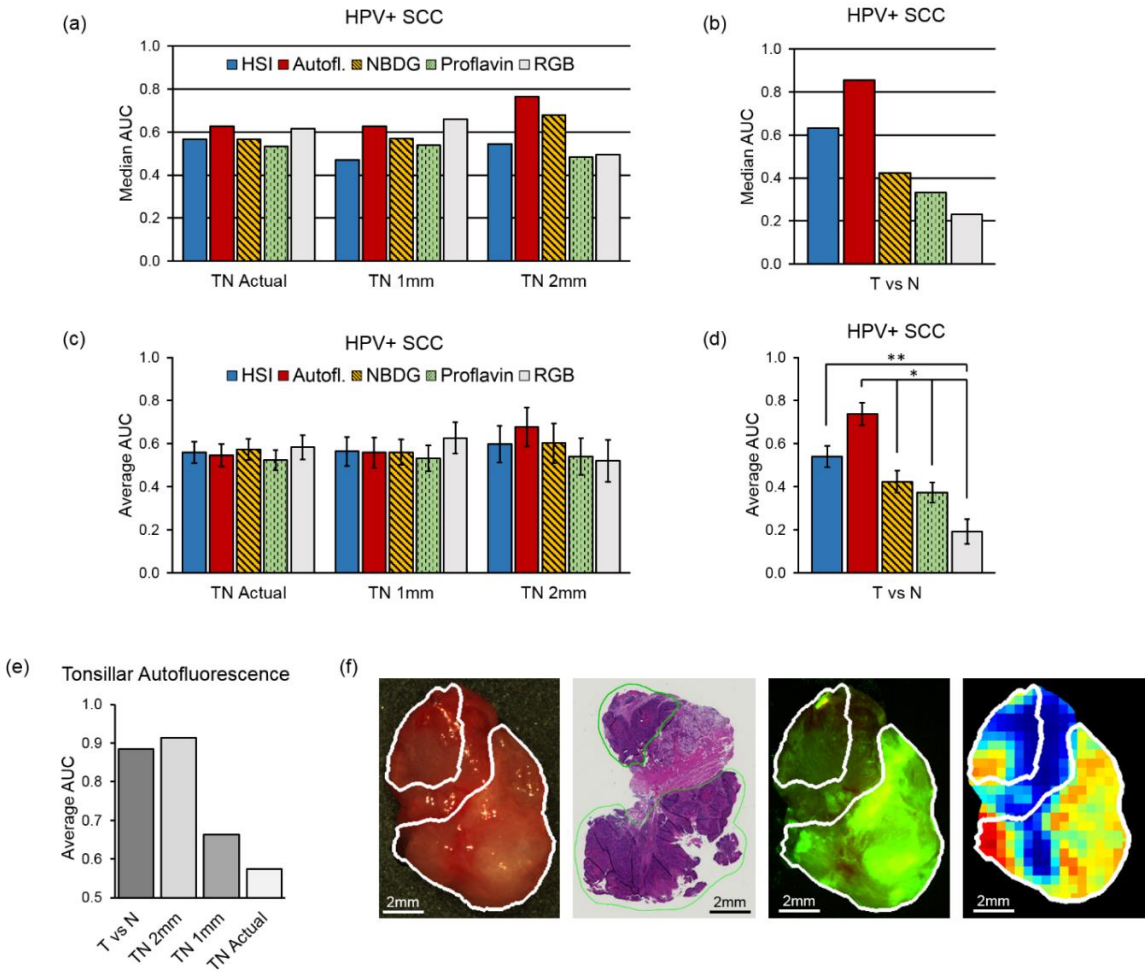

Figure 4-6. Median and average AUC results from inter-patient classification (a value of 0.5 corresponds to random guess). AUC values for the HPV+ SCC cohort: (a) median AUC values TN margin specimens; (b) median AUC values T and N whole specimens; (c) average AUC of TN margin specimens with SEM; (d) average AUC of T and N whole specimens shown with SEM and statistical significance; , shown as (*) for $p<0.05$ and (**) for $p<0.01$; (e) Average AUCs of HPV+ SCC in tonsillar tissues; (f) Representative patient example of HPV+ SCC in tonsillar tissue from the oropharynx. From left to right: RGB made from HSI, histology ground truth, and predicted cancer heat-map. The white and green contours outline the SCC area.

4.4.4 Inter-patient Results of Tumor-Involved Cancer Margins

The results of the inter-patient experiments with HSI in the conventional SCC cohort (N = 88) revealed that testing on specimens of the tumor-only (T) or normal-only

(N) yielded median and average AUCs greater than testing at the tumor margin (TN) at distances up to 2 mm from the margin, 0.92 and 0.87 compared to 0.85 and 0.77, respectively (Figure 4-5a–d). As can been seen in Figure 4-5b, for all distances HSI and autofluorescence outperform 2-NBDG, proflavin, and RGB imaging in average AUC score. Both HSI and autofluorescence significantly outperformed proflavin dye imaging at distances 1 mm (both $p < 0.05$) and 2 mm (both $p < 0.01$) from the cancer margin. HSI and autofluorescence significantly outperform RGB imaging in AUC at the actual cancer margin and 1 mm from the margin (all $p < 0.05$). Table 4-3 shows the full results from conventional SCC with HSI at distances from the cancer margin.

Table 4-3. Performance results from the best label-free HSI methods from the inter-patient experiments for each patient cohort by distance from margin (shown with ± SEM). In the conventional SCC cohort, reflectance-based HSI is presented. Additionally, the HSI results are separated by anatomical location. For the HPV+ SCC cohort, autofluorescence is presented, which all came from the pharynx.

Cohort / Method	Median AUC	Average AUC	Accuracy	Sensitivity	Specificity
Conventional SCC Reflectance-based HSI					
TN Actual	0.75	0.68 ± 0.02	60 ± 2%	60 ± 4%	54 ± 3%
TN 1mm	0.81	0.73 ± 0.02	66 ± 2%	60 ± 4%	61 ± 4%
TN 2 mm	0.85	0.77 ± 0.03	64 ± 3%	65 ± 4%	59 ± 5%
Conventional SCC Reflectance-based HSI					
Oral Cavity, TN 2mm	0.81	0.79 ± 0.04	63 ± 5%	71 ± 8%	49 ± 8%
Tongue, TN 2mm	0.78	0.64 ± 0.07	61 ± 7%	57 ± 9%	53 ± 9%
Nasal Cavity, TN 2mm	0.98	0.93 ± 0.06	79 ± 11%	69 ± 17%	73 ± 24%
Max. Sinus, TN 2mm	0.95	0.78 ± 0.18	58 ± 19%	93 ± 5%	52 ± 18%
Larynx, TN 2mm	0.86	0.85 ± 0.05	79 ± 5%	69 ± 11%	71 ± 9%
Hypopharynx, TN 2mm	0.84	0.78 ± 0.13	42 ± 9%	20 ± 14%	99 ± 1%
Oropharynx, TN 2mm	0.95	0.95 ± 0.001	95 ± 4%	49 ± 49%	78 ± 22%
HPV+ SCC Autofluorescence					
TN Actual	0.63	0.55 ± 0.05	60 ± 4%	45 ± 6%	65 ± 6%
TN 1mm	0.63	0.56 ± 0.07	65 ± 5%	49 ± 8%	62 ± 8%
TN 2 mm	0.77	0.68 ± 0.09	63 ± 7%	64 ± 8%	60 ± 9%

The conventional SCC cohort was separated into the anatomical sites described in Table 4-1, and the highest average AUCs at 2mm from the cancer margin were observed in the nasal cavity (0.93), larynx (0.85), and oropharynx (0.95), while specimens from the tongue performed the lowest (Figure 4-5e). Representative tissue specimens from the oral cavity, nasal cavity, and larynx are classified with HSI and shown in Figure 4-5f–h.

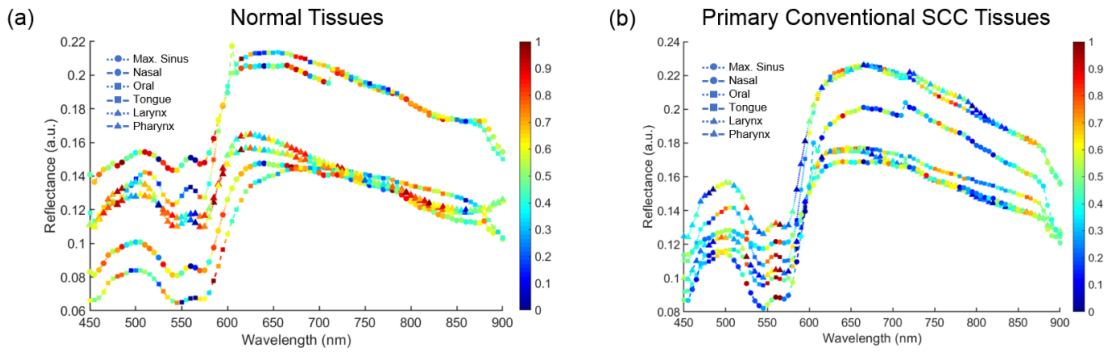

Figure 4-7. Spectral feature saliency from CNN gradients of correctly classified HSI (per grad-CAM) for conventional SCC and normal upper aerodigestive tract tissues separated by anatomical location. (a): Spectral signatures with spectral feature importance for normal tissues from the maxillary sinus, nasal cavity, oral cavity, tongue, larynx, and pharynx. Color of symbol represents the relative importance of the spectral feature for assigning the correct label of the class (0 – blue, low saliency; 1 – red, high saliency). (b): Spectral signatures with spectral feature importance for primary conventional SCC tissues from the maxillary sinus, nasal cavity, oral cavity, tongue, larynx, and pharynx.

For the HPV+ SCC cohort (N = 14) inter-patient experiments, autofluorescence imaging yielded the best results compared to the other imaging modalities at 2 mm from the cancer margin with 0.68 average AUC and 0.77 median AUC (Figure 4-6a,c; not significant, $p > 0.05$). For tonsillar HPV+ SCC tissues (N = 3), the average SCC detection with autofluorescence imaging at 2 mm from the cancer margin was 0.91 AUC (Figure 4-6e). A representative tissue specimen from HPV+ SCC is classified with autofluorescence imaging in Figure 4-6f. Table 4-3 shows the full results from HPV+ SCC with autofluorescence imaging at distances from the cancer margin.

4.5 Discussion

The results of this large study of 293 tissue specimens from 102 patients with SCC show that label-free, reflectance-based hyperspectral imaging and autofluorescence imaging both outperform the fluorescent dye-based imaging methods, *i.e.*, proflavin and 2-NBDG, and this technology could aid in the detection of SCC. The fluorescent dyes employed are not specific enough to target SCC with a high signal-to-noise ratio in ex-vivo tissue specimens because of the large inter-patient variability. Proflavin allows visualization of nuclear structures, but is washed out by excessive keratin. The regional metabolic uptake of 2-NBDG to localize cancerous areas was not evident or demonstrated by the results of this ex-vivo study. Label-free HSI techniques may yield potential but the best machine learning protocols for training HSI classifier is undetermined. It may be task specific, but the results of this study show that with a large SCC HSI database, deep learning algorithms can be trained with high fidelity to work across a large number of anatomical sites in the upper aerodigestive tract.

IPC analysis with frozen sections remains the current standard for intraoperative guidance, but it is time and labor intensive. Across all 102 patients with SCC recruited for this study, an average number of 2.1 IPCs were performed per surgery, each taking about 41 minutes in total. On average, each surgery typically investigated 3.4 tissues, each of which take about 25 minutes to report final diagnosis. The average imaging time for HSI was about 1 minute with up to 35 seconds for HSI classification using the CNN, which is significantly less than IPC.

Detection of SCC for surgical purposes is a challenging task, whether performed by a surgeon, pathologist, or computer-aided optical imaging modality. In the literature, the accuracy of detecting positive or close margins in frozen sections ranges from 71 to 92% with sensitivity from 34 to 77% [203-207]. As sampling and tissue preparation is the main source of error, careful sectioning of small biopsies and vigilant communication is recommended to reduce errors during IPCs [196]. Nonetheless, significant need for guidance remains, with up to 20% to 30% of cases reported with close or positive margin results after SCC resections [203-207]. To this end, to put the SCC detection ability of HSI-based methods into context, we present the pathologist assistant accuracy of 88% for research purposes-only tissue identification. Since current practice is imperfect, the potential benefit of HSI for SCC detection should be evaluated on two criteria: firstly, to establish no potential harm; and secondly, to assess HSI-based intraoperative information that has clinical utility in achieving negative margins, especially considering the time advantage.

The results presented in this study using 293 specimens from 102 patients can be compared to previous pilot studies from our group. Lu et al. 2017 reported results from a small (N = 24) proof-of-concept study using manual ROIs that showed that training and testing on the same patient with HSI yielded an intra-patient accuracy of 89–94% and intra-patient AUC of about 0.96 [33]. Our objective and systematic approach yielded equivalent results using nearly double the intra-patient cohort (N = 47) for distances 1 mm beyond the cancer margin, 85–90% accuracy for conventional SCC and 88–97% accuracy for HPV+ SCC, across all anatomical tissue sites. Moreover, the AUCs obtained from 0.82 to 0.91 for conventional and HPV+ SCC cohorts at 2.25 mm from the cancer margin, importantly,

are not limited by the selection of manual ROIs and include specular glare pixels. Therefore, slightly lower results are to be expected, but provide a more realistic performance estimate for HSI-based methods in the operating room. The experimental results of this study were a median AUC of 0.92 for HSI and 0.93 for autofluorescence for all conventional SCC T vs N tissues using the most patient data in an HSI study to date. The previous proof-of-concept work by Lu et al. reported an accuracy of 85% for T versus N tissues only, an accuracy of 76% for manual ROIs near the cancer-margin, and an overall average AUC of 0.88 for all tissues (T, TN, and N) for SCC at comparable tissues and anatomical sites. In comparison, in this study, for larynx, nasal cavity, and oropharyngeal SCC, we achieved AUC scores of 0.85, 0.93 and 0.95 with accuracies above 79%.

The optical imaging modalities in this study were all acquired using HSI technology, including proflavin, 2-NBDG, and autofluorescence, and all were saved as hypercubes for CNN training. Moreover, even RGB images were generated from HSI, and recent work has suggested that CNNs can recover the full HSI spectrum from RGB composites constructed from HSI [219]. Therefore, it is possible that the results from these modalities benefitted from being HS data, which is one possible explanation for not observing more statistically significant differences. The results of this study are promising at tissue sites that perform with high AUCs in both SCC cohorts. However, the results suggest that HPV+ SCC requires more data to perform well with deep learning. Therefore, the results of this study support the hypothesis that label-free HSI methods significantly outperform the dye-based methods and could provide value for clinical SCC detection.

4.6 Conclusion

In summary, this chapter investigated HSI for SCC detection using all 293 specimens from 102 patients with SCC in the dataset acquired for this thesis, and it was demonstrated that HSI may offer utility for intraoperative SCC detection. This study was published as the first comprehensive report on a large SCC dataset that could be used to train a deep learning model that can predict SCC across multiple anatomical locations with high fidelity. There is a critical need to provide rapid information in the operating room for guidance during SCC resection, with errors reported up to 10–30% of missed positive and close margins in current practice. Our results show that AUCs upwards of 0.80 to 0.90 may be obtained for SCC detection with HSI-based methods in less than 2 minutes for SCC detection, which could save significant time as compared to intraoperative frozen section analysis.

CHAPTER 5. THYROID & SALIVARY TUMOR DETECTION IN 82 PATIENTS WITH HSI

This chapter details the work of using HSI for thyroid and salivary gland tumor detection on all such tissues collected in our dataset. The performance of hyperspectral imaging (HSI) for tumor detection is investigated in ex-vivo specimens from the thyroid (N=200) and salivary glands (N=16) from 82 patients. Tissues were imaged with HSI in broadband reflectance and autofluorescence modes. For comparison, the tissues were imaged with two fluorescent dyes. Additionally, HSI was used to synthesize three-band RGB multiplex images to represent the human-eye response and Gaussian RGBs, which are referred to as HSI-synthesized RGB images. Using histological ground truths, deep learning algorithms were developed for tumor detection. For the classification of thyroid tumors, HSI-synthesized RGB images achieved the best performance with an AUC score of 0.90. In salivary glands, HSI had the best performance with 0.92 AUC score. This study demonstrates that HSI could aid surgeons and pathologists in detecting tumors of the thyroid and salivary glands.

5.1 Introduction

Thyroid cancer incidence has significantly increased worldwide from 1970 to 2012, despite the fact that mortality from thyroid cancer has decreased [220]. Surgery is the standard treatment for thyroid cancers, and the 5-year survival rate for localized or regional thyroid cancers (excluding anaplastic variant) is above 90% [221]. The most common malignant tumor of the thyroid is papillary thyroid carcinoma (PTC), comprising 70% of

thyroid cancers, and there are several variants of PTC, including conventional, follicular, tall-cell, and oncocytic [222]. The initial diagnosis of thyroid tumors is with fine-needle aspiration (FNA) biopsy and histological evaluation of the specimen [222]. Follicular tumors are another cytological type of thyroid neoplasms, which include follicular adenoma, a benign tumor, and follicular thyroid carcinoma (FTC), the malignant form. The requisite diagnostic criterion for follicular carcinoma versus adenoma is definitive invasiveness; no cytological features can provide the diagnosis of FTC, so FNA is therefore useless is making the distinction [223]. Medullary thyroid carcinoma (MTC) is a rare form of thyroid cancer, comprising only 4% of thyroid cancers, that occurs sporadically in most cases, but can be associated with a familial germline mutation [224].

During thyroid tumor resections, intraoperative frozen section (FS) analysis and pathologist consultation can be useful for determining extent of the disease and, according to recent American guidelines, may occasionally confirm malignancy and escalate treatment from partial to total thyroidectomy [225]. For example, in thyroid tumors, 15-30% of preoperative FNA biopsies may be indeterminate [226]. It remains controversial in thyroid tumor surgeries whether the practice of intraoperative FS can provide relevant diagnostic information, as it can be prone to misdiagnosis [226]. In 4% cases with benign intraoperative FS reports, clinically significant malignancy was found, compared to 6.8% in cases where no FS was performed [227]. This translates to a sensitivity of 22% for identifying malignancy in patients with benign FNA [227]. In the literature, it is suggested that the practice of intraoperative FS may lead to over or under treatment of thyroid tumors [226].

Salivary tumors involve the salivary glands, which are a system of exocrine glands in the mouth that produce saliva to initiate digestion. The major salivary glands are the parotid, the submandibular, and the sublingual salivary glands [228]. The classification of benign and malignant salivary tumors is complex, with over 20 distinct entities according to the most recent standard proposed by the World Health Organization [228-230]. Overall, more than 80% of primary tumors of the salivary glands arise in the parotid gland, which is the largest salivary gland [228, 230]. Pleomorphic adenoma is the most common benign tumor of the salivary glands (60%) and typically occurs in the parotid glands [230]. Mucoepidermoid carcinoma is the most common malignant neoplasm of the parotid gland [228]. Adenoid cystic carcinoma is a malignant tumor that can occur with equal likelihood in the submandibular and parotid glands [228]. Polymorphous low-grade adenocarcinoma (PLGA) is a rare malignant tumor, commonly found in minor salivary glands of the hard or soft palate [231]. In surgical resection of salivary tumors, the sensitivity of intraoperative FS for detecting malignant parotid gland tumors with benign FNA was only 33%, suggesting difficulty in diagnosing low grade tumors [232]. Moreover, FS for salivary tumors carries the risk of tumor seeding and may not provide definitive diagnosis [233]. Nonetheless, the combination of preoperative FNA and intraoperative FS leads to high diagnostic accuracy overall for salivary tumors [234].

With the goal of image-guided surgery, hyperspectral imaging (HSI) is an emerging technology in biomedicine that has been used for cancer detection studies both *ex-vivo* and *in-vivo* [14, 211, 212]. HSI has been explored for brain cancer detection *in-vivo* [166, 168]. Additionally, HSI has been proposed for laparoscopic cancer detection in colorectal surgeries with demonstrated potential [22, 129]. The ability of HSI to identify ideal

transection margins for colorectal tissues has been demonstrated after devascularization by blocking vascular anastomoses [235]. Our group reported HSI studies of head and neck cancer using *ex-vivo* human surgical specimens [33, 113, 179, 236].

In order to leave the parathyroid glands intact during surgery, Barberio et al. demonstrated that HSI may be beneficial in detecting parathyroid glands from thyroid tissue during thyroidectomy [237]. For surgeries of salivary tumors, one challenge is leaving the facial nerve intact, which runs through the parotid gland and can cause facial paresis if injured. Wisotzky et al. showed that HSI can identify the facial nerve in the parotid gland [238]. The submandibular and sublingual salivary glands are surrounded by an anatomical variety of normal tissues in the oral cavity. Previous work from our group has demonstrated that HSI can distinguish amongst normal tissues in the oral cavity, such as stratified squamous epithelium, normal salivary gland, and skeletal muscle [179].

In this large study of 82 patients, we perform tumor detection in 200 thyroid tissue specimens from 76 patients in inter-patient testing experiments, and salivary gland tumor detection was investigated using 16 salivary gland tissue specimens from 6 patients. This is the most comprehensive study to date of tumor detection in thyroid and salivary glands to thoroughly assess the feasibility of label-free, non-contact, and non-ionizing HSI-based imaging modalities for computer aided tumor detection. The outcomes of this work will help guide future HSI and autofluorescence studies and determine the specific benefits that HSI may offer for tumor detection in thyroid and salivary gland tissues.

5.2 Methods

In this study, ex-vivo tissue specimens from the thyroid and salivary glands were imaged with optical imaging modalities; histological sections were prepared from the specimens for ground truths; patients were categorized and used to train, validate, and test deep learning algorithms; and performance was calculated to compare the methods.

5.2.1 Ex-vivo Surgical Specimen Dataset

For this study, 216 surgical specimens were acquired from 82 patients undergoing routine resection of thyroid tumors or salivary gland tumors at the Emory University Hospital Midtown, who were recruited by giving written, informed consent to an institutional research coordinator. Table 5-1 shows the categorization of patients and tissue specimens. All patient data were de-identified by the research coordinator. The Institutional Review Board (IRB) of Emory University approved all research protocols and imaging methods. Three types of fresh, ex-vivo surgical specimens were obtained from the surgical pathology department during clinical service. We aimed to acquire a sample of normal tissue (N), tissue from the primary tumor (T), and a specimen of the tumor-involved margin that contains both tumor and normal tissue (TN), all of which were confirmed by histopathological analysis. The size of the tissue specimens was approximately 10×6×2 mm on average. Additionally, the final clinical pathology report was made available after de-identification.

Table 5-1. Number of patients and tissue specimens for this study. Patients are divided by cohort and sub-group. The tissue specimens are categorized into tumor (T), tumor-normal margin (TN), and normal (N).

Group	Patients	Number of Tissue Specimens		
		T	TN	N
Thyroid Tumor Cohort				
Papillary Thyroid Carcinoma (PTC)	54	41	38	59
Medullary Thyroid Carcinoma (MTC) & Insular Carcinoma	6	4	6	8
Follicular Adenoma & Carcinoma	13	14	10	12
Poorly Differentiated Carcinoma	3	4	2	2
Salivary Tumor Cohort				
Parotid Gland Tumor	3	2	3	3
Other Salivary Gland Tumor	3	4	2	2
Total	82	69	61	86

The tissue samples collected for this study were categorized by an experienced pathologist into six groups according to tumor subtype, divided into two broad cohorts: thyroid tumors and salivary gland tumors. The thyroid tumor cohort was comprised of 200 tissue specimens from 76 patients. The malignant tumors included in this cohort were PTC (N=54), MTC (N=5), insular carcinoma (N=1), follicular carcinoma (N=8), and poorly differentiated thyroid carcinoma (N=3). The benign tumors of the thyroid were follicular adenoma (N=5). The only thyroid cohort tissues excluded from this study were six patients with benign thyroid hyperplasia/goiter.

The cohort of salivary gland tumors was comprised of 16 tissue specimens from 6 patients. Two patients had benign pleomorphic adenoma (N=2) of the parotid gland. Four patients had malignant tumors of the salivary glands: mucoepidermoid carcinoma (N=1), salivary duct carcinoma of the parotid gland (N=1), PLGA of the hard palate (N=1), and

adenoid cystic carcinoma (N=1). The patient demographics and relevant cancer properties are shown in Table 5-2.

Table 5-2. Patient demographics and cancer properties for the patients recruited for this study. Intraoperative pathologist consultations (IPCs) using intraoperative frozen section (FS) analysis are also reported.

Property	Number	Percentage
Demographics		
Average Age (years)	49.7	-
Male	28	34%
Female	54	66%
Tobacco History	26	32%
Ethnicity		
White	47	57%
Black	19	24%
Asian	6	7%
Other / Unknown	10	12%
Primary Location / Tumor Type		
Thyroid Gland		
Papillary Carcinoma	54	66%
Follicular Carcinoma	8	10%
Follicular Adenoma	5	6%
Medullary Carcinoma	5	6%
Poorly Differentiated Carcinoma	3	4%
Insular Carcinoma	1	1%
Salivary Gland		
Pleomorphic Adenoma	2	2%
Adenoid Cystic Carcinoma	1	1%
Mucoepidermoid Carcinoma	1	1%
Adenocarcinoma (PLGA)	1	1%
Salivary Duct Carcinoma	1	1%
Cancer Stage		
pT1	19	23%
pT2	19	23%
pT3	26	32%
pT4	9	11%
Avg. T Size (cm)	3	-
N+ Status	33	40%
IPC (Averages)		
Surgeries with IPC	19	23%
IPC/Surgery	1.6	-
Tissues/Surgery	1.1	-
Time/IPC (min)	19.1	-
Time/Tissue (min)	17.3	-

5.2.2 Optical Imaging Modalities

To assess the ability of HSI for tumor detection, several other optical imaging modalities were acquired for comparison, including both label-free and fluorescent dye-based methods. It was hypothesized that HSI would outperform fluorescence methods due to lack of sufficient target specificity in 2-NBDG and proflavin. In the following sections, the image acquisition systems are described for hyperspectral reflectance imaging, HSI-synthesized RGB multiplex imaging, autofluorescence imaging, and two fluorescent dye-based imaging techniques: 2-NBDG and proflavin.

5.2.2.1 Hyperspectral Imaging

A CRi Maestro HS system (Perkin Elmer Inc., Waltham, Massachusetts) was used to acquire HSI of the *ex-vivo* specimens. The HS system performs spectral scanning from 450 to 900 nm using a Xenon light source and liquid crystal tunable filter (LCTF) with 5 nm spectral resolution [33, 239]. The image size of the HSI was 1040×1392×91 pixels (height×width×spectral bands), and the corresponding specimen-level spatial resolution was 25 μm per pixel. Acquisition time for an HSI was approximately one minute.

The raw HS data (I_{raw}) were normalized band-by-band (λ) by subtracting the inherent dark current of the sensor and dividing by a white reference disk for all pixels (x,y), according to the following equation.

$$I_{norm}(x, y, \lambda) = \frac{I_{raw}(x, y, \lambda) - I_{dark\ current}(x, y, \lambda)}{I_{white\ ref}(x, y, \lambda) - I_{dark\ current}(x, y, \lambda)}$$

The average spectral signatures after white-dark calibration are shown for all groups included in this paper by cohort in Figure 5-1.

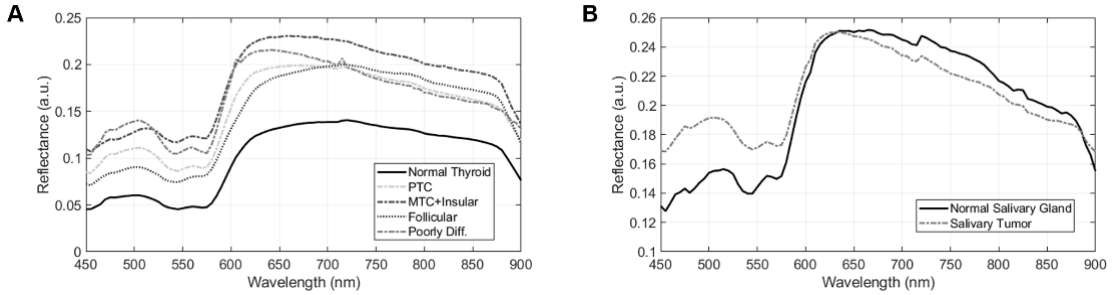

Figure 5-1. Average hyperspectral signatures for the tissues in the thyroid cohort (a) and the salivary cohort (b).

5.2.2.2 HSI-Synthesized RGB Images

A multiplex image is a synthetic composite image generated from a hyperspectral image. For this work, several three-band (RGB) multiplex images were synthesized from the normalized reflectance HSI hypercubes. The first synthetic RGB was generated from the HSI by applying a Gaussian kernel in each color region, which is referred to as HSI-synthesized Gaussian RGB composite. The second RGB image was constructed from human color perception curves originally proposed by Judd et al. 1951 and expanded by Vos 1978 [240]. For some tissues, a standard RGB image was also captured for comparison by an RGB camera. Figure 5-2 shows a representative surgical tissue specimen of thyroid cancer from an RGB-captured camera, HSI-synthesized Gaussian RGB multiplex, and HSI-synthesized RGB with human eye perception.

In this paper, we use HSI to simulate three-band images using RGB multiplex imaging. However, the sensors of RGB cameras typically employ the Bayer filter [241, 242] for adapting RGB color spectrums similar to human-eye. While the spectrum of blue and green are typically consistent, different RGB camera sensor types have differing sensitivity to the red channel components between 400 and 500 nm [241, 242]. Therefore, for a subset of thyroid tumor specimens, the component of red channel response between 400 and 500 nm was manipulated to simulate if this would have an effect on performance from the HSI-synthesized human eye RGB multiplex images.

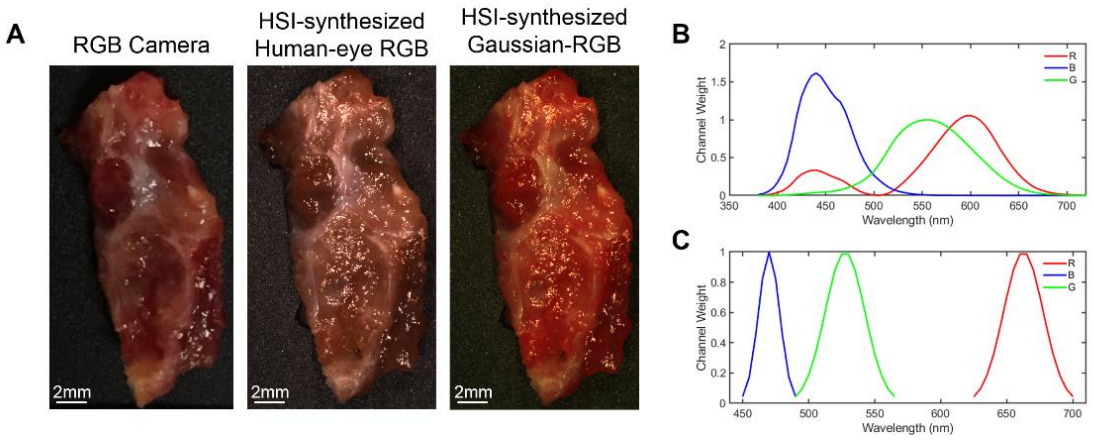

Figure 5-2. A representative specimen of thyroid cancer. (a) Left to right: RGB image from standard RGB camera; HSI-synthesized RGB human-eye multiplex image made from reflectance HSI using Vos et al. 1978 method; HSI-synthesized Gaussian RGB multiplex image made from reflectance HSI. (b) Spectral signatures of human-eye color perception of red (R), green (G), and blue (B) colors proposed by Vos et al. 1978. (c) Gaussian kernels used for generating Gaussian RGB multiplex images.

5.2.2.3 Autofluorescence imaging

Autofluorescence imaging uses the emission from intrinsic fluorophores in biological tissue that are stimulated to fluoresce by external excitation light at specific wavelengths. The autofluorescence images were captured using a 455 nm excitation source

and a 490 nm long-pass filter using the CRi Maestro imaging system. The long-pass filter removes any of the external light from the source that would be reflected into the image and allows capturing images of emission-only photons, according to Stokes' theorem. The autofluorescence images were acquired from 500 to 720 nm in 10 nm increments to produce a hypercube with 23 spectral bands and final size of 1040×1392×23 pixels.

5.2.2.4 2-NBDG Imaging

A fluorescently tagged glucose molecule, 2-deoxy-2-[(7-nitro-2,1,3-benzoxadiazol-4-yl)amino]-D-glucose (2-NBDG), is a dye that targets cancer regions by producing a stronger signal measured from regions with higher metabolic glucose uptake. After the hyperspectral and autofluorescence imaging methods described above, the tissues were incubated for 20 minutes in a 160 µM 2-NBDG solution (Cayman Chemical, Ann Arbor, MI, USA) at 37 degrees Celsius, quickly rinsed in 1× phosphate buffered solution (PBS) to remove excess dye, and fluorescence imaging was performed using the CRi Maestro. The images were acquired with the same excitation light source at 455 nm and a long-pass filter at 490 nm from 500 to 720 nm in 10 nm increments, producing a hypercube that has 23 spectral bands.

5.2.2.5 Proflavin Imaging

The second dye used for fluorescence imaging was proflavin dye, which is unaffected by previous 2-NBDG dye because it has a significantly stronger optical signal comparatively. Proflavin fluorescent dye binds to DNA and thus allows visualization of nuclear morphology, which can improve the ability of machine learning based cancer detection methods [213]. Keratin is also a target of proflavin dye, but this should not affect

the glandular tissues involved in this study [213, 214]. For proflavin imaging, the tissue samples were incubated for 2 minutes in a 0.01% proflavin solution (Sigma Aldrich, St. Louis, MO, USA) at room temperature, and the tissues were rinsed in PBS before imaging with the CRi Maestro. The images were acquired with an excitation light source at 455 nm and a long-pass filter at 490 nm from 500 to 720 nm in 10 nm increments, producing a hypercube that contains 23 spectral bands.

5.2.3 Histological Ground Truth

The ground truths for the optical imaging modalities were achieved using digitized histology imaging. After acquiring all images, the tissue specimens were inked at the top, bottom, left and right edges, and back surface of the tissue to identify tissue orientation in histological sections. Tissues were then fixed in formalin, paraffin embedded, and sectioned with a microtome, and 5 μm slices were made from the surface that was optically imaged. The first high quality slice was kept to serve as the histology ground truth, processed with hematoxylin and eosin staining, and digitized using whole-slide scanning at 40× objective [243]. A board-certified pathologist with expertise in head and neck pathology annotated the tumor and normal areas on the digital histology images.

A binary mask was made of the contoured digital histology images, which served as the ground truth for the optical imaging modalities. Due to tissue deformations during histological processing and slide preparation, the histology ground truth masks needed to be registered to the gross-level optical images. The digital histology slide was registered in a semi-automated method according to a previously established pipeline of affine, land-mark, and deformable registration to the gross-level HSI [184, 216]. The transformation

was applied to the binary histology mask, and thus a ground truth mask was generated matching the gross-level optical images of the tissue specimens.

5.2.4 Experimental Design

The deep learning experimental designs involved training, validation, and testing. However, for the two cohorts, thyroid tumors (N=76 patients) and salivary tumors (N=6 patients), there were different designs of the data partitioning, which was required because of the significant differences in sample size. Therefore, the thyroid cohort is used to produce fully-independent inter-patient results. However, the salivary cohort is used to produce intra-patient training and testing, as described below in detail.

5.2.4.1 Thyroid Tumors

Tumor detection of the thyroid gland was performed in fully-independent patients, divided across 5 folds. Each fold served as the fully-independent testing group, while training and validation was performed on the patients in the remaining 4 folds, as depicted in Figure 5-3. This design was selected to allow test-level performance metrics for all 76 thyroid patients.

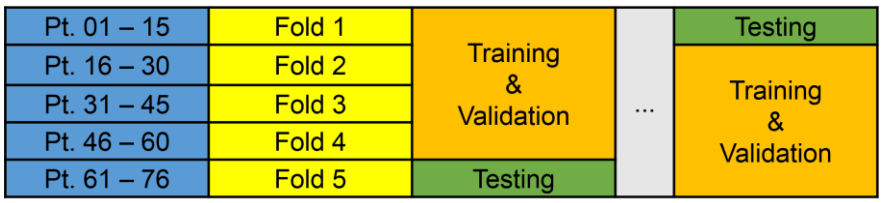

Figure 5-3. Schematic depicting the experimental design of fully-independent training, validation, and testing paradigms for the 76 patient thyroid tumor cohort.

5.2.4.2 Salivary Gland Tumors

Tumor detection in the limited sample size of salivary gland tumors was performed using intra-patient experiments. Training and validation was performed on the patients' primary tumor (T) and normal (N) tissues, and testing was performed on the tumor-normal (TN) margin tissue. Figure 5-4 shows a schematic diagram of the training and testing paradigm. The salivary gland cohort was separated into tumors of the parotid gland (N=3 patients) and other salivary glands (N=3), as shown in Table 5-1.

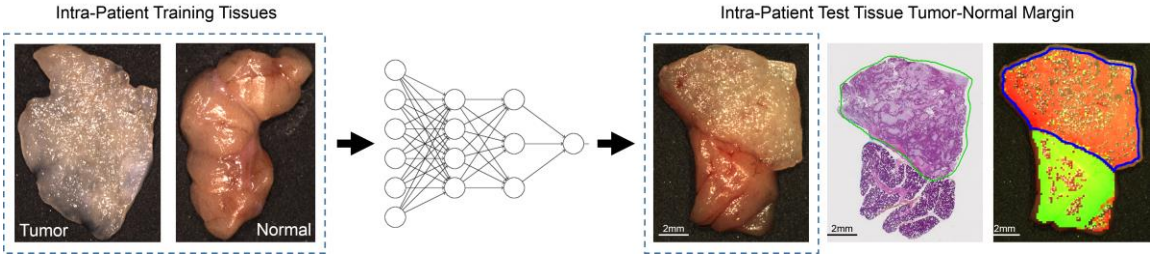

Figure 5-4. Flow diagram of intra-patient experiments of the salivary gland, with representative tumor of the parotid gland. Intra-patient T and N tissues were used for MLP (multilayer perceptron) training, and TN tissue specimens were used for testing. The histological ground truth is shown with tumor contour in green. The predicted tumor heat-map overlay onto the RGB image is shown with tumor predictions (red) and normal predictions (green). Areas of specular glare in the heat-map are not classified, and the ground-truth tumor contour is in blue.

5.2.5 Convolutional Neural Network

For thyroid tumor detection using 200 tissue specimens from 76 patients, a convolutional neural network (CNN) was developed for the effective classification of thyroid tissue into tumor and normal using a patch-based approach. The inception-v4 CNN architecture [217] was selected because it is one of the top performing CNNs on standard tasks like Image-Net, yet has a manageable number of hyperparameters. HSI data has several unique challenges due to data size. Therefore, the CNN required modification for

HS data pre-processed into image-patches of size 25×25×*C* pixels, where *C* represents the number of spectral bands. The first convolutional layers were modified for the smaller patch-size necessitated by HS data, and the operating resolution in the modular inception blocks was reduced to allow more efficient training and classification using the CNN. Additionally, squeeze-and-excitation modules were added to increase the performance of the CNN [244]. The implemented CNN architecture schematic is detailed in Figure 5-5.

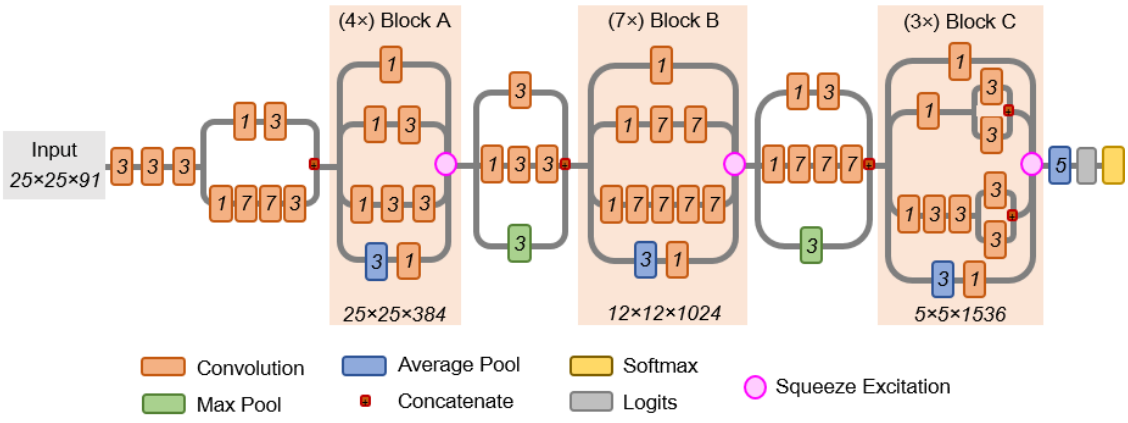

Figure 5-5. Schematic diagram of the modified inception v4 CNN architecture. The CNN was customized to operate on the 25×25×91 patch-size selected. The receptive field size and number of convolutional filters is shown at bottom of each inception block. The convolutional kernel size used for convolutions is shown in italics inside each convolution box. Squeeze-and-excitation modules were added to the CNN to increase performance.

Image patches (25×25×*C*) were generated using a sliding window approach with a stride of 13 pixels, and the data pre-processing of HSI was performed in MATLAB (MathWorks, Inc., Natrick, MA). All deep learning programming was done in the TensorFlow python software package [192] on an Ubuntu machine and accelerated with CUDA execution on Titan-XP NVIDIA GPUs (Nvidia Corp., Santa Clara, CA). The CNN loss function was cross-entropy, the optimizer utilized was Adadelta with an initial learning rate of 1.0, and validation performance was calculated every 2 epochs of training data.

Training each CNN model was performed for 14 epochs of 8× augmented (reflections and rotations) training data, which took about 23 hours to train. Deployment of a fully-trained CNN model on a single GPU to classify a new thyroid tissue specimen, which consisted of hundreds of patches, was 20 ± 8 seconds (avg. ± st. dev.) for all imaging modalities. The heatmaps were produced by averaging the results of overlapping pixel regions in image-patches, since the 25×25 patches were produced with a stride of 13 pixels, which was used to produce a smoother and less coarse final result.

5.2.5.1 Multilayer Perceptron

For salivary gland tumor detection of 16 tissue specimens from 6 patients, a simplified artificial neural network, called a multilayer perceptron (MLP), was used for intra-patient detection with spectral information only. The MLP consisted of a 91 unit spectral vector input, a single hidden layer with 128 neurons, and an output layer of 2 nodes (normal or tumor). This simplified MLP was applied only to the salivary gland tumor cohort and selected to limit overfitting in this small dataset. The salivary gland cohort was separated into parotid gland tumors (N=3 patients) and other salivary gland tumors (N=3 patients).

The spectral signatures of tissue were extracted by local averaging of 5×5 pixel blocks to reduce noise. Image pre-processing was used to remove the specular glare pixels from both training and testing. For each group, the normal (N) and tumor-only (T) specimens' spectra were used for training (85%) and a subset for validation (15%), and the tumor-normal (TN) margin tissue spectra was used for testing. For the parotid group and other salivary gland group separately, all patients' training samples were combined into

one training group (6 tissues), and the three TN test specimens were classified. Training was performed on the order of a few minutes, and testing was produced in about one second.

5.2.6 Performance Evaluation

The principal evaluation metric used for this study was area under the curve (AUC) of the receiver operator curve (ROC). The AUC score was selected because it is robust to class imbalances within tissues and provides an estimate of performance at all possible thresholds of separating the normal and tumor classes. Additionally, the accuracy, sensitivity, and specificity were calculated and reported using the tumor probability threshold from the validation data. All results were calculated on a tissue specimen level and averaged. Additionally, for the final testing results of both cohorts, the 10 pixels at the edge of tissue, corresponding to 0.25 mm, were excluded from performance calculations. The imaging protocol for ex-vivo tissue specimens was performed using a flat imaging surface, so the tissue edges created unnatural curvature where the tissue was too thin to provide an adequate imaging signal. Statistical significance was calculated for the test results using Student's t-test and a 0.05 *p*-value threshold.

5.3 Results

5.3.1 Thyroid Tumor Detection

Tumor detection for the thyroid cohort with all cancer types combined (N=76) demonstrated that the HSI-synthesized RGB multiplex images generated from the HSI were the best performing results in terms of average AUC scores with 0.89 and 0.90 for

HSI-synthesized Gaussian RGB multiplex and the HSI-synthesized human-eye RGB multiplex, respectively. HSI performed with an average AUC score of 0.86. Full results of thyroid tumor detection by AUC score, accuracy, sensitivity, and specificity for all imaging methods are shown in Table 5-3 and separated by cancer type.

Table 5-3. Performance results of the optical imaging modalities for the thyroid tumor cohort (average ± SEM). The best performing modality for each groups' evaluation metrics is bolded.

		AUC	Accuracy	Sensitivity	Specificity
HSI	All Thyroid Tumors (*N*=82)	0.86 ± 0.02	78 ± 2%	80 ± 3%	74 ± 3%
	PTC (*N*=54)	0.86 ± 0.02	76 ± 2%	79 ± 3%	71 ± 4%
	MTC & Insular Ca. (*N*=6)	0.81 ± 0.09	85 ± 5%	85 ± 5%	72 ± 11%
	Follicular Ad. & Ca. (*N*=13)	0.90 ± 0.04	80 ± 4%	80 ± 7%	82 ± 5%
	Poorly Diff. Ca. (*N*=3)	0.90 ± 0.08	81 ± 15%	73 ± 23%	96 ± 4%
Auto-fluorescence	All Thyroid Tumors (*N*=82)	0.85 ± 0.02	76 ± 2%	**83 ± 2%**	68 ± 3%
	PTC (*N*=54)	0.81 ± 0.03	72 ± 2%	79 ± 3%	62 ± 4%
	MTC & Insular Ca. (*N*=6)	0.86 ± 0.06	80 ± 5%	83 ± 7%	78 ± 6%
	Follicular Ad. & Ca. (*N*=13)	**0.95 ± 0.02**	**87 ± 3%**	**93 ± 3%**	81 ± 5%
	Poorly Diff. Ca. (*N*=3)	0.98 ± 0.01	**95 ± 1%**	**95 ± 3%**	93 ± 1%
2-NBDG Dye	All Thyroid Tumors (*N*=82)	0.86 ± 0.02	78 ± 2%	75 ± 3%	80 ± 3%
	PTC (*N*=54)	0.84 ± 0.02	76 ± 2%	72 ± 3%	79 ± 3%
	MTC & Insular Ca. (*N*=6)	0.84 ± 0.05	80 ± 5%	74 ± 8%	72 ± 10%
	Follicular Ad. & Ca. (*N*=13)	0.92 ± 0.03	84 ± 4%	82 ± 6%	89 ± 4%
	Poorly Diff. Ca. (*N*=3)	0.91 ± 0.08	79 ± 11%	81 ± 15%	89 ± 11%
Proflavin Dye	All Thyroid Tumors (*N*=82)	0.83 ± 0.02	75 ± 2%	68 ± 3%	80 ± 3%
	PTC (*N*=54)	0.82 ± 0.02	73 ± 2%	66 ± 4%	77 ± 4%
	MTC & Insular Ca. (*N*=6)	0.70 ± 0.08	71 ± 4%	60 ± 9%	71 ± 8%
	Follicular Ad. & Ca. (*N*=13)	0.91 ± 0.03	81 ± 4%	73 ± 7%	**93 ± 2%**
	Poorly Diff. Ca. (*N*=3)	0.98 ± 0.01	95 ± 2%	91 ± 5%	**97 ± 2%**
HSI-synthesized Gaussian-RGB	All Thyroid Tumors (*N*=82)	0.89 ± 0.02	79 ± 2%	77 ± 2%	**82 ± 3%**
	PTC (*N*=54)	0.87 ± 0.02	77 ± 2%	76 ± 3%	**79 ± 4%**
	MTC & Insular Ca. (*N*=6)	0.95 ± 0.03	88 ± 4%	91 ± 4%	82 ± 8%
	Follicular Ad. & Ca. (*N*=13)	0.90 ± 0.02	77 ± 3%	67 ± 6%	91 ± 2%
	Poorly Diff. Ca. (*N*=3)	0.98 ± 0.01	94 ± 3%	92 ± 4%	95 ± 4%
HSI-synthesized Human-Eye RGB	All Thyroid Tumors (*N*=82)	**0.90 ± 0.02**	**79 ± 2%**	80 ± 2%	79 ± 3%
	PTC (*N*=54)	**0.88 ± 0.02**	**78 ± 2%**	**80 ± 3%**	76 ± 4%
	MTC & Insular Ca. (*N*=6)	**0.96 ± 0.02**	**88 ± 3%**	**93 ± 3%**	**85 ± 6%**
	Follicular Ad. & Ca. (*N*=13)	0.92 ± 0.02	76 ± 3%	68 ± 7%	86 ± 4%
	Poorly Diff. Ca. (*N*=3)	**0.99 ± 0.01**	91 ± 3%	92 ± 3%	93 ± 5%

The average and median AUC scores are presented in Figure 5-6A and B across all thyroid tumor types with statistical significance. Combining all thyroid tumors, both

implementations of three-band, HSI-synthesized RGB multiplex imaging (average AUC score of 0.89 for Gaussian-RGB multiplex and 0.90 for human-eye RGB multiplex) outperformed autofluorescence (0.85 AUC score), 2-NBDG (AUC score of 0.86), and proflavin (0.83 AUC score) to a degree of statistical significance (all $p<0.05$). Additionally, HSI-synthesized human-eye RGB multiplexing also significantly outperformed HSI ($p<0.05$). For the PTC sub-group ($N=54$), both HSI-synthesized RGB multiplex images statistically outperformed autofluorescence, 2-NBDG, and proflavin (all $p<0.05$). For the MTC group ($N=6$), both HSI-synthesized RGB multiplex images significantly outperformed 2-NBDG and proflavin (all $p<0.05$). For the follicular tumor group, autofluorescence outperforms the other methods in AUC score, but the difference is not significant ($p>0.05$). Lastly, poorly differentiated thyroid carcinomas were classified with the highest AUC score from HSI-synthesized human-eye RGB multiplex imaging, but not significantly ($p>0.05$).

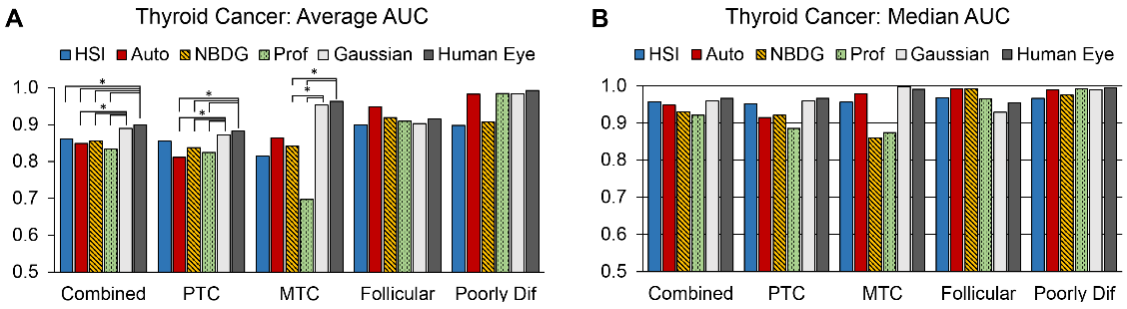

Figure 5-6. Average and median AUC scores from thyroid tumor detection. (a) average AUC scores for thyroid tumor detection across all tissue specimens grouped by tumor subtype; statistical significance (*, $p<0.05$) is shown above. (b) median AUC scores of tumor subtype detection.

The different imaging modalities and respective probability heat-maps for tumor detection are shown in Figure 5-7 for all groups of thyroid tumors. As can be seen, HSI

shows the most consistent heat-maps around regions of specular glare, compared to the HSI-synthesized RGB multiplex methods. As shown in Figure 5-6B, the median AUC scores show that HSI (0.95) and the two HSI-synthesized RGB multiplex methods (0.95 and 0.96) have approximately equivalent median performance for combined thyroid tumors.

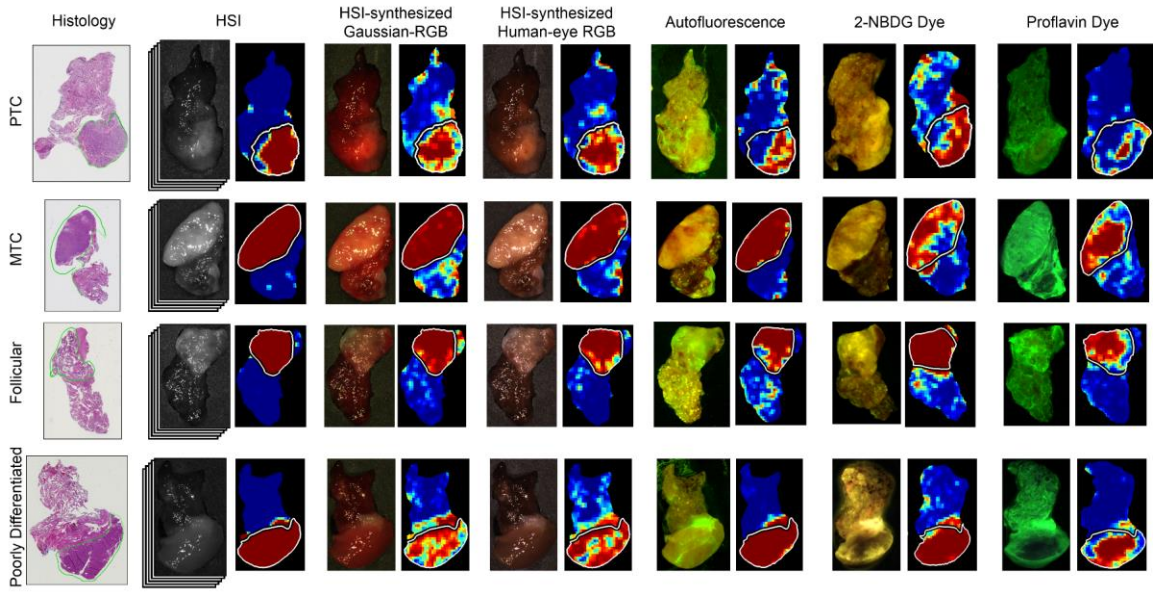

Figure 5-7. Representative tissue images and corresponding classification heat-maps from all modalities from patients with thyroid carcinoma. Columns from left to right: histology, HSI with heat-map, HSI-synthesized Gaussian-RGB multiplex with heat-map, HSI-synthesized human-eye RGB multiplex with heat-map, autofluorescence with heat-map, 2-NBDG dye image with heat-map, Proflavin dye image with heat-map. Rows from top to bottom: papillary thyroid carcinoma (PTC), medullary thyroid carcinoma (MTC), follicular thyroid carcinoma, and poorly differentiated thyroid carcinoma. The contours in white (in heat-maps) and green (on histology) outline the cancerous regions. Predicted tumor heat-maps range from dark blue (predicted normal) to dark red (predicted cancer).

The median AUC scores are substantially greater than the averages, which indicates that the distribution of performance results tends to be more accurate than the average reflects. Histogram analysis of percent difference of the HSI and HSI-synthesized RGB multiplex imaging methods shows that HSI-synthesized RGB outperforms HSI in a

relatively small number of tissue specimens, which causes the average AUC scores for HSI-synthesized RGB multiplexing to be greater than HSI. Figure 5-8A and C show the histograms of percent difference in tissues from HSI to both Gaussian-RGB multiplex and the human-eye RGB multiplex, respectively. Additionally, Figure 5-8B and D show the tissue specimens with the largest differences in AUC scores compared to HSI, where HSI-synthesized RGB multiplexing still works quite well.

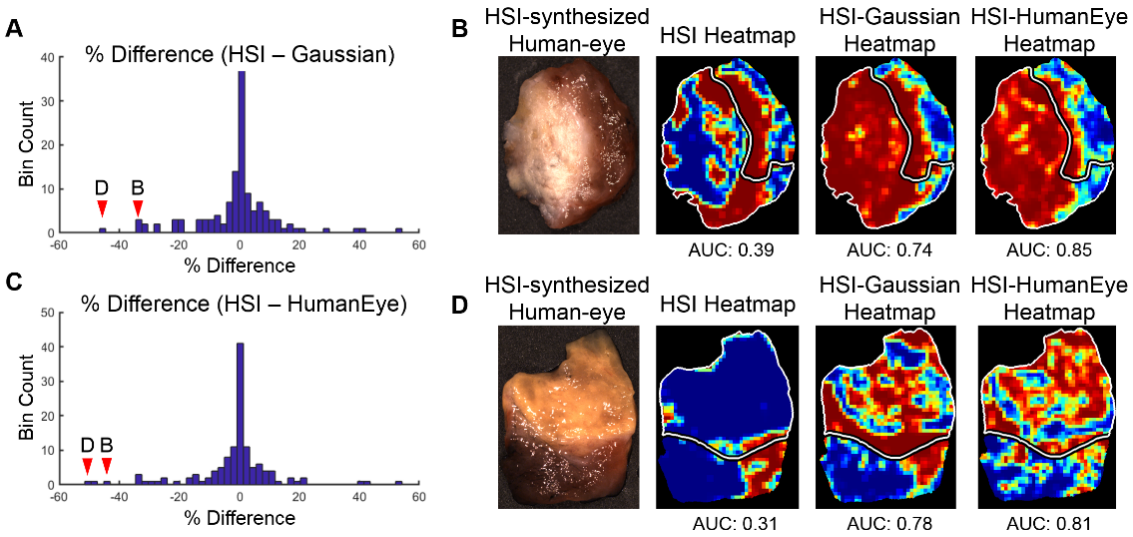

Figure 5-8. Differences in AUC score performance comparing HSI against HSI-synthesized Gaussian-RGB multiplex and HSI-synthesized human-eye RGB multiplexing. (a) Histogram of percent difference in AUC scores of tissue specimens between HSI and HSI-synthesized Gaussian-RGB multiplex imaging. The arrows show the bins that contain the patient specimens shown in the right panels, which are the two worst performing tissues. (b) RGB image of tissue specimen with large difference in AUC score performance between heat-maps produced from HSI, HSI-synthesized Gaussian multiplex, and HSI-synthesized human-eye multiplex image. (c) Histogram of percent difference in AUC scores of tissue specimens between HSI and HSI-synthesized human-eye RGB multiplex imaging. The arrows show the bins that contain the patient specimens shown in the right panels, which are the two worst performing tissues. (d) RGB image of the tissue specimen with the largest difference in AUC score performance between heat-maps produced from HSI, HSI-synthesized Gaussian multiplex, and HSI-synthesized human-eye multiplex image. The tumor margin is delineated in white.

The three-band HSI-synthesized RGB multiplex images from HSI are meant to represent RGB imaging. However, these multiplex images are still constructed from HSI data. Standard RGB camera sensors have different responses to the component of the red channel between 400 and 500 nm. Therefore, for fold 1 of the HSI-synthesized human-eye RGB multiplex thyroid tumor detection, the red channel component from 400 to 500 nm was multiplied by half and by zero, and two more CNNs were trained. The results are plotted in Figure 5-9. The original human-eye multiplex result for fold 1 using the original 400-500 nm red component was an AUC score of 0.90 for thyroid tumor detection. Completely eliminating the 400-500 nm red component (by multiplying by zero) in human-eye multiplex still resulted in an equivalent AUC score of 0.90 for thyroid tumor detection (p>0.05). Lastly, an equivalent AUC score of 0.89 for thyroid tumor detection was obtained when the 400-500nm red component was set to half the original value for human eye multiplexing (p>0.05).

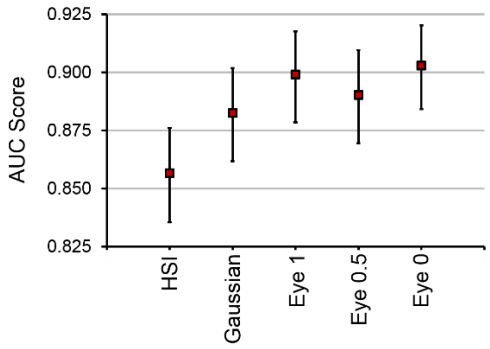

Figure 5-9. AUC score results with 95% confidence interval from 1 fold of the testing data comparing different methods of constructing HSI-synthesized RGB multiplex images. Left to right: Standard HSI; Gaussian-RGB from HSI; original HSI-synthesized human-eye RGB from HSI; human-eye RGB from HSI synthesized with half of the 400-500 nm red component; human-eye RGB from HSI synthesized with zero of the 400-500 nm red component.

To further investigate how CNN methods can utilize HSI and the relevant wavelengths for correctly predicting normal and thyroid tissues, we incorporated the gradient class-activated maps (grad-CAM) algorithm [218]. Briefly, the method is used for tracing the most relevant gradients from the input data to the class of interest, either normal or tumor, which is used to infer spectral feature saliency. The mean spectral signatures and class-activated gradients were averaged for 89 tissues that were correctly classified with high AUC scores and separated into normal thyroid and tumor (Figure 5-10). As can be observed in Figure 5-10A, the salient spectral features for correctly classifying normal thyroid tissues were from 570-700 nm. In Figure 5-10B, the most salient spectral features for correctly classifying thyroid tumors were also in the range of 550-700 nm, with additional bands near 500 and 750 contributing some lesser importance for classification.

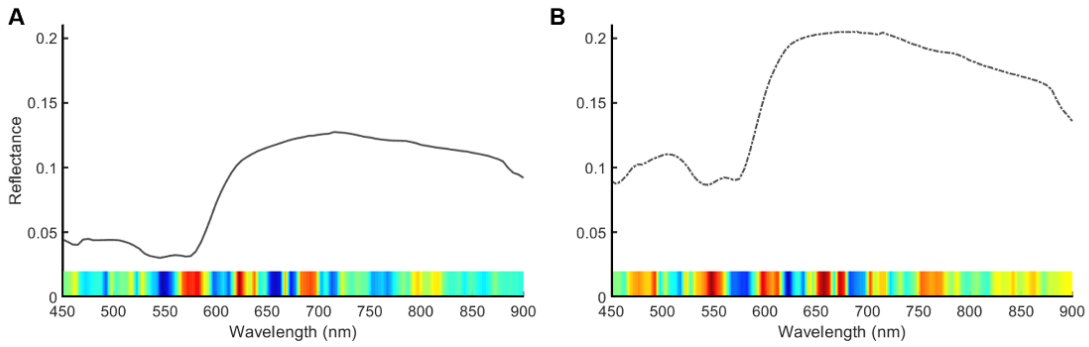

Figure 5-10. Mean spectral signatures of correctly-classified normal thyroid tissues (a) and thyroid tumors (b). The saliency of spectral features is identified below each plot using the grad-CAM technique. Red hues represent the most important features for correctly predicting each tissue class, and blue color hues represent less important wavelengths for correctly predicting each class.

5.3.2 Salivary Gland Tumor Detection

The intra-patient tumor detection results of the salivary gland tumors cohort are separated into parotid gland tumors and other salivary gland tumors. For parotid gland tumors, HSI was the best performing imaging modality with an AUC score of 0.92, accuracy of 88%, sensitivity of 90%, and specificity of 79% (all differences were not significant, $p>0.05$). For tumors of other salivary glands, autofluorescence was the best performing imaging modality with an AUC score of 0.80, accuracy of 84%, sensitivity of 77%, and specificity of 85% (all differences were not significant, $p>0.05$). The full results are shown in Table 5-4.

Table 5-4. Performance results from all optical imaging modalities for the salivary tumor cohort, separated by salivary gland group (average ± SEM). The best performing modality for each groups' evaluation metrics is bolded.

Group	Imaging	AUC	Accuracy	Sensitivity	Specificity
Parotid (*N*=3)	HSI	**0.92 ± 0.04**	88 ± 4%	90 ± 4%	**79 ± 7%**
	Autofluorescence	0.60 ± 0.30	87 ± 2%	**99 ± 1%**	26 ± 26%
	2-NBDG Dye	0.77 ± 0.12	87 ± 2%	97 ± 3%	31 ± 31%
	Proflavin Dye	0.64 ± 0.32	**89 ± 2%**	96 ± 2%	54 ± 27%
	Gaussian-RGB	0.66 ± 0.15	81 ± 5%	97 ± 2%	38 ± 24%
	Human-Eye	0.72 ± 0.02	78 ± 15%	95 ± 5%	35 ± 20%
Other Salivary (*N*=3)	HSI	0.59 ± 0.14	76 ± 12%	55 ± 29%	85 ± 15%
	Autofluorescence	**0.80 ± 0.14**	**84 ± 10%**	**77 ± 13%**	85 ± 6%
	2-NBDG Dye	0.52 ± 0.08	76 ± 11%	49 ± 28%	84 ± 16%
	Proflavin Dye	0.62 ± 0.10	77 ±11%	60 ± 30%	78 ± 22%
	Gaussian-RGB	0.45 ± 0.03	70 ± 10%	62 ± 31%	57 ± 43%
	Human-Eye	0.78 ± 0.15	84 ± 9%	76 ± 12%	**87 ± 6%**

5.4 Discussion

The results of this extensive study suggest that label-free HS-based imaging and autofluorescence does indeed outperform the two fluorescent dye-based imaging methods

tested here for thyroid tumor detection, but not to a significant degree. Interestingly, we discovered that HSI-synthesized RGB multiplex imaging significantly outperforms all imaging methods tested for thyroid tumor detection, including HSI and autofluorescence (p<0.05). For salivary tumor detection, HSI performs best in the parotid gland and autofluorescence in performs best in other salivary glands, but no difference was significant. As can be observed in Figure 5-6B, one main conclusion from this work is that with sufficiently large datasets, many different optical imaging modalities can be used to create deep learning algorithms for tumor detection with median AUC scores of 0.90 and upwards. This phenomenon can be seen specifically for the thyroid tumors combined cohort.

The experiments for the thyroid cohort and salivary gland cohort were processed separately in different ways because of vastly different numbers of tissue samples collected. The thyroid tumor cohort experiment was investigated in fully-independent testing patients because 200 tissue specimens from 76 patients were available. The salivary tumor cohort was comprised of only 16 tissues from 6 patients, so intra-patient experiments were performed using tumor-only and normal specimens for training and testing on tumor-normal margin tissues.

The hypothesis that HSI-based methods would outperform fluorescent dye-based methods was upheld in the thyroid tumor combined category, largely because it was supported in the PTC group (N=54), which comprises 71% of cases. However, it was not supported for MTC, FTC, and poorly differentiated thyroid carcinoma groups (all p values were not significant). Additionally, the thyroid tumor detection results show that HSI-synthesized human-eye RGB multiplex imaging made from HSI statistically outperforms

reflectance-based HSI. Despite differences in average AUC scores, the median values are equivalent around 0.95 (see Figure 5-6). Exploring this phenomenon further in Figure 5-8, it was demonstrated that only a few tissues differ between the HSI and HSI-synthesized RGB multiplex modalities. Moreover, the probability heat-maps from HSI seem to provide more consistent classification around regions of significant specular glare compared to the HSI-synthesized multiplex methods (Figure 5-7). These results are consistent with a previous study from our group that was limited to only 11 thyroid patients, in which RGB composite images (AUC score of 0.95) also outperformed HSI (AUC score of 0.92) [33].

The purpose of the three-band HSI-synthesized multiplex images synthesized from HSI was to represent standard RGB imaging from a standard camera. However, these multiplex images are still constructed from HSI data. Additionally, there are differences in the spectral responses to the red channel component from 400-500 nm in standard RGB camera sensors. The impact of this red component value was studied, and no effect was observed in AUC score by altering these values for HSI-synthesized human-eye RGB multiplexing. To provide physical intuition for this conclusion, the grad-CAM method reveals that the most salient spectral features for correctly classifying normal thyroid tissues were from 570-700 nm, well above this range. Future studies are required to investigate if a standard RGB camera would indeed outperform HSI directly. Additionally, future work is needed to capture more thyroid tumor HSI data with higher spatial and spectral resolution HS cameras. It is possible that the spectral resolution of 5 nm in this LCTF spectral-scanning HS system was inadequate for this study.

5.5 Conclusion

In conclusion, this chapter presents the extensive study using 216 glandular tissue samples from all 82 patients with thyroid and salivary gland tumors in our acquired dataset for this thesis work to evaluate the performance of HSI for tumor detection of the thyroid and salivary glands. For comparison to HSI, the tissues were imaged with label-free autofluorescence and two fluorescent dyes, 2-NBDG and proflavin dye. Additionally, HSI-synthesized three-band multiplex images, representing the human-eye response and Gaussian RGBs, were synthesized from HSI. Several CNNs were developed for tumor detection that perform with median AUC scores of 0.90 and higher for all imaging modalities in combined thyroid tumors. Investigating each group specifically, our results suggest that HSI-synthesized human-eye RGB multiplexing can classify thyroid tumors significantly better than HSI. In salivary glands, label-free HSI and autofluorescence may offer the best performance for tumor detection. The conclusions from this chapter demonstrate that HSI may aid surgeons and pathologists in tumor detection in glands.

CHAPTER 6. TISSUE SUB-COMPONENT CLASSIFICATION USING HSI

This chapter details the investigation of the ability for HSI to perform sub-component analysis, beyond normal versus cancer, in a subset of the head and neck tissues. We developed a method to use HSI, a non-contact optical imaging modality, and convolutional neural networks (CNNs) to perform an optical biopsy of ex-vivo, surgical gross-tissue specimens, collected from 21 patients undergoing surgical cancer resection. Using a cross-validation paradigm with data from different patients, normal tissue from the upper aerodigestive tract can be sub-classified into squamous epithelium, muscle, and gland with an average AUC score of 0.94. After separately training on thyroid tissues, the CNN can discriminate between thyroid carcinoma and normal thyroid with an AUC score of 0.95 and 92% accuracy/sensitivity/specificity. Moreover, further analysis was performed, detecting medullary thyroid carcinoma from benign multi-nodular goiter (MNG) with an AUC of 0.93, and papillary thyroid carcinoma from MNG with an AUC of 0.91. These preliminary results demonstrate that an HSI-based optical biopsy method using CNNs can provide multi-category diagnostic information for normal and cancerous head and neck tissue.

6.1 Introduction

Cancers of the head and neck are the sixth most common cancer world-wide, including cancers that are predominately of squamous cell origin, for instance the oral cavity, nasopharynx, pharynx, and larynx, and others like carcinomas of the thyroid gland

[139]. Major risk factors include consumption of tobacco and alcohol, exposure to radiation, and infection with human papilloma virus (HPV) [140, 245]. Approximately 90% of cancer at sites including the lips, gums, mouth, hard and soft palate, and anterior two-thirds of the tongue are squamous cell carcinoma (SCC) [2]. The diagnostic procedure of SCC typically involves physical examination and surgical evaluation by a physician, tissue biopsy, and diagnostic imaging, such as PET, MRI, or CT. Patients with SCC tend to present with advanced disease, with about 66% presenting as stage III or IV disease, which requires more procedures for successful treatment of the patient [141]. The standard treatment for these cancers usually involves surgical cancer resection with potential adjuvant therapy, such as chemotherapy or radiation, depending on the extent, stage, and location of the lesion. Successful surgical cancer resection is a mainstay treatment of these cancers in order to prevent local disease recurrence and promote disease free survival [4].

Previous studies have investigated the optical properties of normal and malignant tissues from areas of the upper aerodigestive tract [246-249]. Muller et al. acquired *in-vivo* reflectance-based spectroscopy from normal, dysplasia, inflammation, and cancer sites in the upper aerodigestive tract from patients with SCC to extract tissue parameters that yield biochemical or structural information for identifying disease. Varying degrees of disease and normal tissue could be distinguished because of their different optical properties that were believed to be related to collagen and nicotinamide adenine dinucleotide (NADH) [246]. Similarly, Beumer et al. acquired reflectance spectroscopy measurements from 450 to 600 nm from patients with SCC and implemented an inverse Monte Carlo method to derive oxygenation-based tissue properties from the optical signatures, which were found to be significantly different in malignant and non-malignant tissues [249].

Hyperspectral imaging (HSI) is a non-contact, optical imaging modality capable of acquiring a series of images at multiple discrete wavelengths, typically on the order of hundreds of spectral bands. Preliminary research demonstrates that HSI has potential for providing diagnostic information for a myriad of diseases, including anemia, hypoxia, cancer detection, skin lesion and ulcer identification, urinary stone analysis, enhanced endoscopy, and many potential others in development [14-25]. Supervised machine learning and artificial intelligence algorithms have demonstrated the ability to classify images after being allowed to learn features from training or example images. One such method, convolutional neural networks (CNNs), have demonstrated astounding performance at image classification tasks due to their capacity for robust handling of training sample variance and ability to extract features from large training data sizes [26, 27].

The need for an imaging modality that can perform diagnostic prediction could potentially aid surgeons with real-time guidance during intra-operative cancer resection. This study aims to investigate the ability of HSI to classify tissues from the thyroid and upper aerodigestive tract using convolutional neural networks. This work was initially presented as a conference proceedings and oral presentation [171]. First, a simple binary classification is performed, i.e. cancer versus normal, and second, multi-class sub-classification of normal upper aerodigestive tract samples is investigated. If proven to be reliable and generalizable, this method could help provide intra-operative diagnostic information beyond palpation and visual inspection to the surgeon's resources, perhaps enabling surgeons to achieve more accurate cuts and biopsies, or as a computer-aided diagnostic tool for physicians diagnosing and treating these types of cancer.

6.2 Methods

To investigate the ability of HSI to perform optical biopsy, we recruited patients with thyroid or upper aerodigestive tract cancers into our study, acquired and processed gross-level HSI of freshly excised tissue specimens, trained our convolutional neural network, and evaluated system performance.

6.2.1 Experimental Design

In collaboration with the Otolaryngology Department and the Department of Pathology and Laboratory Medicine at Emory University Hospital Midtown, 21 head and neck cancer patients who were electing to undergo surgical cancer resection were recruited for our study to evaluate the efficacy of using HSI for optical biopsy [147, 170]. From these 21 patients, a total of 63 excised tissue samples were collected. From each patient, 3 tissue samples were obtained from the primary cancer gross specimen in the surgical pathology department after the primary cancer had been resected. The specimens were selected to include tumor, normal, and a tissue at the tumor-normal interface. Each specimen was typically around 10 by 10 mm in area and 3 mm in depth. The collected tissues were kept in cold PBS during transport to the imaging laboratory where the specimens were scanned with a hyperspectral imaging system [33, 113].

Two regions of interest were used for this study: firstly, the upper aerodigestive tract sites, including tongue, larynx, pharynx, and mandible; and secondly, the thyroid and associated carcinomas. Head and neck squamous cell carcinoma (HNSCC) of the aerodigestive tract represented the first group, comprised of 7 patients. Normal tissue was obtained from all patients in the HNSCC group, and SCC was obtained from 6 of these

patients. In head and neck cancers, the non-cancerous tissues adjacent to SCC may be dysplastic, inflammatory, or keratinized, which could affect the hyperspectral imaging results. Therefore, the normal tissues included in this study were regions of normal tissue that were not dysplastic or heavily inflamed tissues. The thyroid group consisted of 14 patients total and included one benign neoplasm and three malignant neoplasms: benign multi-nodular goiter (MNG, 3 patients), classical-type papillary thyroid carcinoma (cPTC, 4 patients), follicular-type papillary thyroid carcinoma (fPTC, 4 patients), and medullary thyroid carcinoma (MTC, 3 patients), respectively.

After imaging with HSI, tissues were fixed in formalin, underwent haemotoxylin and eosin (H&E) staining, paraffin embedded, sectioned, and digitized. A certified pathologist with head and neck expertise confirmed the diagnoses of the *ex-vivo* tissues using the digitized histology slides in Aperio ImageScope (Leica Biosystems Inc, Buffalo Grove, IL, USA). The histological images serve as the ground truth for the experiment.

6.2.2 Hyperspectral Imaging and Processing

The hyperspectral images were acquired using a CRI Maestro imaging system (Perkin Elmer Inc., Waltham, Massachusetts), which is comprised of a Xenon white-light illumination source, a liquid crystal tunable filter, and a 16-bit charge-coupled device (CCD) camera capturing images at a resolution of 1040 by 1,392 pixels and a spatial resolution of 25 μm per pixel [113, 145, 170, 183]. The hypercube contains 91 spectral bands, ranging from 450 to 900 nm with a 5 nm spectral sampling interval. The average imaging time for acquiring a single HSI was about 1 minute.

The hyperspectral data were normalized at each wavelength (λ) sampled for all pixels (*i*, *j*) by subtracting the inherent dark current (captured by imaging with a closed camera shutter) and dividing by a white reference disk according to the following equation [145, 170]. The normalized HSI spectral signatures are plotted by class in Figure 6-1.

$$I_{norm}(\lambda, i, j) = \frac{I_{raw}(\lambda, i, j) - I_{dark\ current}(\lambda, i, j)}{I_{white\ ref}(\lambda, i, j) - I_{dark\ current}(\lambda, i, j)}$$

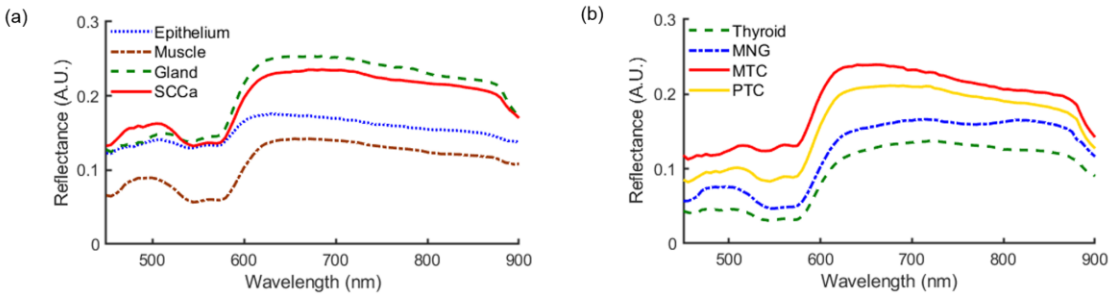

Figure 6-1. Normalized spectral signatures, which were averaged across tissues. Presented by anatomical location: (a) normal tissue and SCC of the upper aerodigestive tract, (b) normal, benign, and carcinoma of the thyroid.

Specular glare is created on the tissue surfaces due to wet surfaces completely reflecting incident light. Glare pixels do not contain useful spectral information for tissue classification and are hence removed from each HSI by converting the RGB composite image of the hypercube to grayscale and experimentally setting an intensity threshold that sufficiently removes the glare pixels, assessed by visual inspection.

A schematic of the classification scheme is shown in Figure 6-2. For binary cancer classification, the classes used are normal aerodigestive tissue versus SCC, and medullary and papillary thyroid carcinoma versus normal thyroid tissue. For multi-class classification of oral and aerodigestive tract tissue, squamous epithelium, skeletal muscle, and salivary

glands in the oral mucosa are used. For binary classifications of thyroid cancer, classical-type papillary thyroid carcinoma, medullary thyroid carcinoma, and multi-nodular thyroid goiter tissue are used.

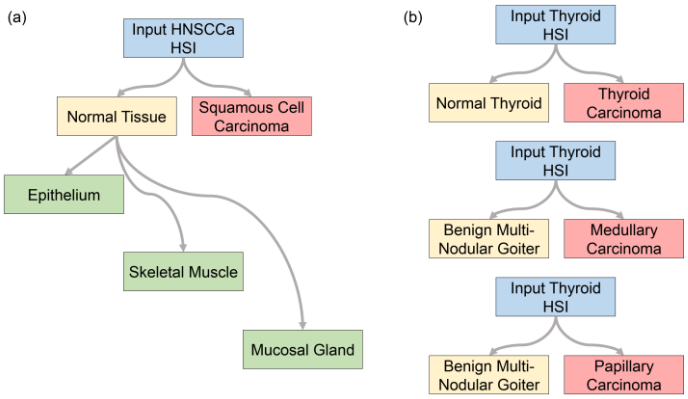

Figure 6-2. Tissue classification scheme. (a) For classification of the HNSCC group, first a binary classification is considered to test the ability of the classifier to distinguish normal samples from SCC samples. Next, histologically confirmed normal samples are sub-classified squamous epithelium, skeletal muscle, and mucosal salivary glands. (b) For classification of the thyroid group, first a binary classification is considered to test the ability of the classifier to distinguish normal thyroid samples from thyroid carcinoma of multiple types. In addition, thyroid HSI classification is tested to discriminate MNG from MTC and to discriminate MNG from classical-type PTC.

To avoid introducing error from registration of tissue samples that contain a tumor-normal boundary, only samples that contain exactly one class were used for binary classification. For example, the tumor sample and normal sample are held out for testing, so that validation performance can be evaluated on both class types. Out of the initial 63 samples acquired from 21 patients recruited for this study, this elimination process excluded 22 tissue samples because it was found that two tumor-normal margin samples were obtained from one patient. The normal samples from the patients with MNG thyroid neoplasm were not included in the binary cancer detection experiment. Additionally, after clinical histological evaluation, it was determined that one *ex-vivo* specimen from one

papillary thyroid carcinoma patient was an adenomatoid nodule. This type of lesion is reported in the literature to commonly cause misdiagnoses in initial needle biopsies [250, 251], and importantly, this lesion type was not adequately represented in the current study. Therefore, the two remaining tissue samples from this patient were removed from this study because we aimed to have balanced tissue specimens from the thyroid carcinoma patients, normal and tumor. After all exclusionary criteria were determined, there were 36 tissue specimens from 20 patients, 7 HNSCC and 13 thyroid, incorporated in this study, and the epithelium, muscle, and gland tissue components were selected as ROIs from 7 tissue samples from the HNSCC group. The number and type of tissue specimens is detailed in Table 6-1.

Table 6-1. Number of ex-vivo tissue specimens included in this study from the 13 patients with thyroid neoplasms and 7 patients with SCC. The number of image patches for CNN classification obtained from each specimen type is also reported.

Group	Class	No. Tissue Specimens	Total Patches
Thyroid	Normal Thyroid	10	14,491
	MNG	3	9,778
	MTC	3	10,334
	Classical PTC	3	6,836
	Follicular PTC	4	13,200
HNSCC	Squamous Epithelium	4	6,366
	Skeletal Muscle	3	5,238
	Mucosal Gland	4	5,316
	SCC	6	4,008

For training and testing the CNN, each patient HSI needs to be divided into patches. Patches are produced from each HSI after normalization and glare removal to create 25×25×91 non-overlapping patches that do not include any “black-holes” where pixels

have been removed due to specular glare; see Table 6-1 for the total number of patches per class. Glare pixels are intentionally removed from the training dataset to avoid learning from impure samples. In addition, patches were augmented by 90, 180, and 270 degree rotations and vertical and horizontal reflections, to produce six times the number of samples. For cancer classification, the patches were extracted from the whole tissue. While for multi-class sub-classification of normal tissues, the regions of interest comprised of the classes of target tissue were extracted using the outlined gold-standard histopathology images.

6.2.3 Convolutional Neural Network

The convolutional neural networks used in this study were built from scratch using the TensorFlow application program interface (API) for Python. A high-performance computer was used for running the experiments, operating on Linux Ubuntu 16.04 with 2 Intel Xeon 2.6 GHz processors, 512 GB of RAM, and 8 NVIDIA GeForce Titan XP GPUs. Two distinct CNN architectures were implemented for HNSCC classification, which incorporated inception modules, and thyroid classification, which used a 3D architecture. Both architectures are detailed below and shown schematically in Figure 6-3. During the following experiments, only the learning-related hyper-parameters that were adjusted between experiments, which include learning rate, decay of the Adadelta gradient optimizer, and batch-size. Within each experiment type, the same learning rate, rho, and epsilon were used, but some cross-validation iterations used different numbers of training steps because of earlier or later training convergence.

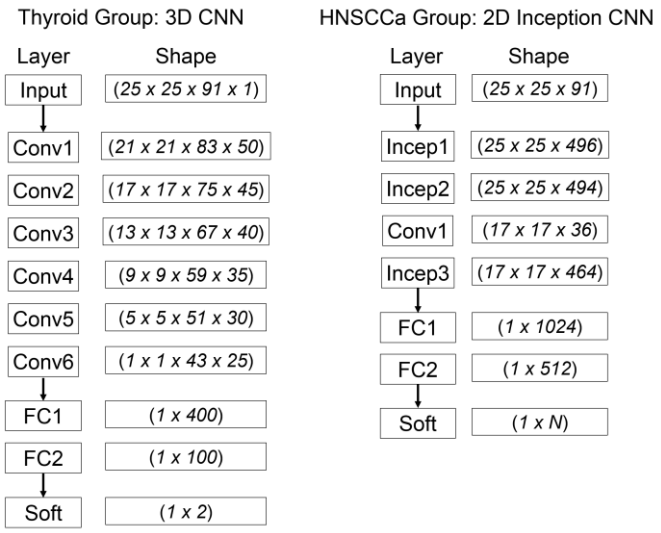

Figure 6-3. CNN architectures implemented for classification of HSI of thyroid tissue (left) and tissue from the upper aerodigestive tract (right).

To classify thyroid tissues, a 3D-CNN based on AlexNet, an architecture originally designed for ImageNet classification, was implemented using TensorFlow [26, 192]. The model consisted of six convolutional layers with 50, 45, 40, 35, 30, and 25 convolutional filters, respectively. Convolutions were performed with a convolutional kernel of 5×5×9, which correspond to the *x*-*y*-λ dimensions. Following the convolutional layers were two fully connected layers of 400 and 100 neurons each. A drop-out rate of 80% was applied after each layer. Convolutional units were activated using rectified linear units (ReLu) with Xavier convolutional initializer and a 0.1 constant initial neuron bias [252]. Step-wise training was done in batches of 10 patches for each step. Every one thousand steps the validation performance was evaluated, and the training data were randomly shuffled for improved training. Training was done using the AdaDelta, adaptive learning, optimizer for reducing the cross-entropy loss with an epsilon of 1×10^{-8} and rho of 0.9 [253]. For thyroid normal versus carcinoma, the training was performed at a learning rate of 0.1 for two to six thousand steps depending on the iteration. For MNG versus MTC and for MNG versus

cPTC, the training was done at a learning rate of 0.005 for exactly two thousand steps for all iteration.

Classification of upper aerodigestive tract tissues was hypothesized to be a more complex task, so a 2D-CNN architecture was constructed to include a modified version of the inception module appropriate for HSI that does not include max-pools and uses larger convolutional kernels, implemented using TensorFlow [26, 27, 192]. The modified inception module simultaneously performs a series of convolutions with different kernel sizes: a 1×1 convolution; and convolutions with 3×3, 5×5, and 7×7 kernels following a 1×1 convolution. The model consisted of two consecutive inception modules, followed by a traditional convolutional layer with a 9×9 kernel, followed by a final inception module. After the convolutional layers were two consecutive fully connected layers, followed by a final soft-max layer equal to the number of classes. A drop-out rate of 60% was applied after each layer. For binary classification, the number of convolutional filters were 355, 350, 75, and 350, and the fully connected layers had 256 and 218 neurons. For multi-class classification, the number of convolutional filters were 496, 464, 36, and 464, and the fully connected layers had 1024 and 512 neurons. Convolutional units were activated using rectified linear units (ReLu) with Xavier convolutional initializer and a 0.1 constant initial neuron bias [252]. Step-wise training was done in batches of 10 (for binary) or 15 (for multi-class) patches for each step. Every one thousand steps the validation performance was evaluated and the training data were randomly shuffled for improved training. Training was done using the AdaDelta, adaptive learning, optimizer for reducing the cross-entropy loss with an epsilon of 1×10^{-8} (for binary) or 1×10^{-9} (for multi-class) and rho of 0.8 (for binary) or 0.95 (for multi-class) [253]. For normal oral tissue versus SCC binary

classification, the training was done at a learning rate of 0.05 for five to fifteen thousand steps depending on the patient-held-out iteration. For multi-class sub-classification of normal aerodigestive tract tissues, the training was done at a learning rate of 0.01 for three to five thousand steps depending on the patient-held-out iteration.

6.2.4 Validation

The final layer of the CNN labels each test case as the class with the highest probability overall, so each test patch has exactly one label. In addition, the probabilities of each test patch belonging to all classes are output from the network. All testing is done on a patch-based level to ensure class accuracy. The typical classification time for one patient's entire HSI was on the order of several minutes. The class probabilities for all patches of a test patient case are used to construct receiver operator characteristic (ROC) curves using MATLAB (MathWorks Inc, Natick, MA, USA). For binary classification, only one ROC curve is created per patient test case, but for multi-class classification, each class is used to generate a respective ROC curve; true positive rate and false positive rate are calculated as that class against all others. The CNN classification performance was evaluated using leave-one-patient-out external-validation to calculate the sensitivity, specificity, and accuracy, defined below, using the optimal operating point of each patient's ROC curve [33, 170]. The area under the curve (AUC) for the ROC curves is calculated as well and averaged across patients.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Samples}}$$

6.3 Results

6.3.1 Aerodigestive Tract Normal Tissue Sub-classification

Using a leave-one-patient-out cross-validation paradigm with HSI obtained from different patients, the CNN can distinguish SCC from normal oral tissues with an AUC of 0.82, 81% accuracy, 81% sensitivity, and 80% specificity. Table 6-2 shows the full results. The ROC curves for all HNSCC patients are shown in Figure 6-4. Additionally, normal oral tissues can be sub-classified into squamous epithelium, muscle, and glandular mucosa using a separately trained CNN, with an average AUC of 0.94, 90% accuracy, 93% sensitivity, and 89% specificity. Representative normal sub-classification results are shown in Figure 6-5, and full results are detailed in Table 6-3.

Table 6-2. Results of inter-patient CNN classification (leave-one-patient-out cross-validation). Values reported are averages across all patients shown with standard deviation.

Group	Classification	# Tissue Specimens	AUC	Acc. (%)	Sens. (%)	Spec. (%)
Thyroid	Normal vs Carcinoma	20	0.95±0.07	92±9	92±8	92±10
	cPTC vs MNG	7	0.91±0.10	86±13	86±14	86±13
	MTC vs MNG	6	0.93±0.04	87±5	88±4	85±7
HNSCC	Normal vs SCC	13	0.82±0.13	81±11	81±15	80±16
	Multi-class	7	0.94±0.08	90±9	93±6	89±13

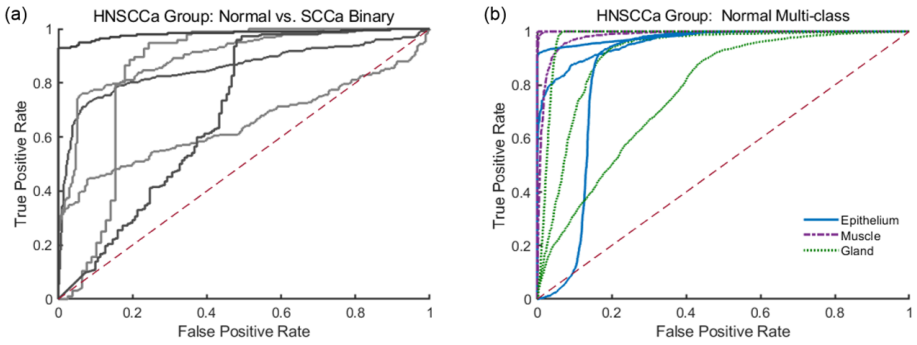

Figure 6-4. HSI classification results ROC curves for HNSCC experiments generated using leave-one-out cross-validation. (a) binary classification of SCC and normal head and neck tissue; (b) multi-class sub-classification of normal aerodigestive tract tissues.

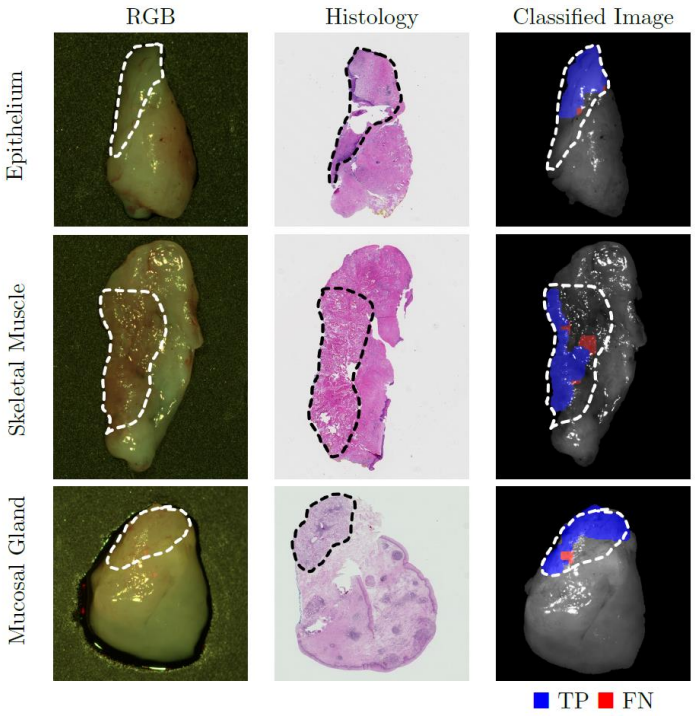

Figure 6-5. Representative results of sub-classification of normal oral tissues. Left: HSI-RGB composites are shown with ROI of the tissue type outlined. Center: Respective histological gold standard with corresponding ROI outlined. Right: Artificially colored CNN classification results of the ROI only. True positive results representing correct tissue-subtype are visualized in blue, and false negatives are shown in red. Tissue within the ROI that is shown in grayscale represents tissue that is not classified due to glare pixels or insufficient area to produce the necessary patch-size.

Table 6-3. Results of inter-patient CNN classification of sub-classified normal upper aerodigestive tract tissues. Values reported are averages shown with standard deviation.

Group	No. Tissue Specimens	AUC	Accuracy	Sensitivity	Specificity
Squamous Epithelium	4	0.94±0.06	90±5%	91±3%	91±7%
Skeletal Muscle	3	0.99±0.01	98±3%	98±3%	97±4%
Mucosal Gland	4	0.89±0.10	83±13%	90±8%	79±18%

6.3.2 Thyroid Tissue Sub-classification

After separately training on thyroid tissue, the CNN differentiates between thyroid carcinoma and normal thyroid with an AUC of 0.95, 92% accuracy, 92% sensitivity, and 92% specificity. The ROC curves for all thyroid patients are shown in Figure 6-6. Moreover, the CNN can discriminate MTC from benign MNG with an AUC of 0.93, 87% accuracy, 88% sensitivity, and 85% specificity. Classical-type papillary thyroid carcinoma is differentiated from MNG with an AUC of 0.91, 86% accuracy, 86% sensitivity, and 86% specificity.

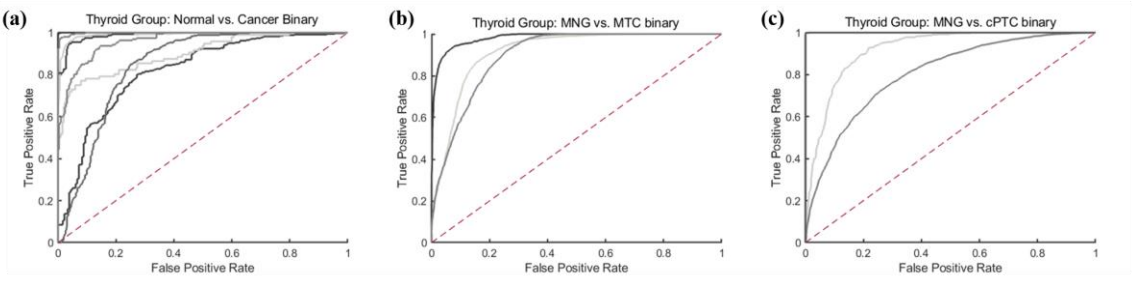

Figure 6-6. HSI classification results ROC curves for HNSCC experiments generated using leave-one-out cross-validation. (c) binary classification of normal thyroid and thyroid carcinomas; (d) binary classification of MNG and MTC; (e) binary classification of MNG and classical PTC.

6.4 Discussion

We developed a deep learning based classification method for hyperspectral images of fresh surgical specimens. The study demonstrated the ability of HSI and convolutional neural networks for discriminating between normal tissue and carcinoma. The novel results of normal tissue sub-classification into categories of squamous epithelium, skeletal muscle, and glandular mucosa demonstrate that there is further classification potential for HSI.

A review of surgical cases found that head and neck surgery has the most intra-operative pathologist consultations (IPC) per surgery, typically around two consultations per surgery, compared to other organ systems [254]. The average time at our institution was about 45 minutes per IPC. Despite being currently un-optimized and performed off-line, our method takes around 5 minutes including imaging time, classification, and post-processing. The main benefit is that the proposed method does not require excising tissue or any tissue processing to provide diagnostic information of the surgical area. Additionally, our method is demonstrated to be significantly faster than an average IPC. However, we do not suggest that the proposed method could replace an IPC, but rather provide guidance during surgery to reduce the number or increase the quality of IPCs.

In this study, the limited patient dataset reduces the generalizability of the results. In addition, the ROI technique for outlining tissues of interest for normal multi-class sub-classification creates the potential to introduce error into the experiment. Both of these issues could be resolved by utilizing a large number of patient data. Moreover, the gross tissue specimens utilized in this study are entirely cancer or normal, so the cancer samples are comprised of sheets of malignant cells. In order for the proposed method to be extended

to the operating room to aid in the resection of cancer, the method needs to be investigated on detecting cancer cells extending beyond the cancer edge. Therefore, future studies will investigate the ability of the proposed method to accurately predict the ideal resection margin of cancerous tissues.

When the diagnosis of thyroid cancer is suspected, a needle biopsy is performed, which can provide information on the histological and intrinsic type of cancer that is present. It is not uncommon for thyroid cancer to be present with benign hyperplasia of the non-malignant portions of the effected lobe. However, the co-occurrence of both MTC and PTC is rare, although cases have been documented [255-257]. We hypothesized that HSI has further potential than binary cancer detection, as we explored in this work, and that different types of cancer can be identified from benign hyperplasia. Therefore, we performed a set of two binary classifications (MTC versus MNG, and PTC versus MNG) to show both can be successfully identified from MNG. In order to determine if HSI can detect normal thyroid and the range of thyroid neoplasms: MTC, PTC, hyperplasia, follicular carcinoma, adenomas, and nodules; all in one multi-class approach, more patient data collection and more investigation into the robustness HSI and the classifiers needs to be performed.

We acknowledge that with a limited patient set for the experiments detailed, we do not have a fully-independent test set and employ a leave-one-patient-out cross validation approach, as was reported in the results section. This approach has several limitations with regards to potential over-fitting from hyper-parameter tuning. However, in order to avoid bias, the CNN architectures and number of filters in each network were not adjusted during the experiments conducted. As stated, a more complex CNN design was used for SCC

because it was deemed a more complex problem. During the experiments, the only parameters that were adjusted were the learning-related parameters, which are detailed in the method section. These included the learning rate and the epsilon and rho of Adadelta, which control the decay of learning rate and gradient optimization. Importantly, the same learning rate, rho, and epsilon were used within each experiment type so that all cross-validation iterations had the same hyper-parameters. We found that different experiments, for example HNSCC binary compared to multi-class, required different learning-based parameters because they trained at different rates.

Another potential source of errors from the cross-validation experiments was the effect of over-fitting during training of each cross-validation iteration, within an experiment type. With the relatively small sample sizes employed in this study, swapping one patient from the validation to training set could drastically change the training time of the network. Therefore, performance of each cross-validation iteration was evaluated every 1,000 training steps, and different training steps for cross-validations were sometimes used. In the methods section, we report the range of training steps for each experiment type; in order to reduce bias, the same training step number was used when it was possible for all or most cross-validation iterations.

Another limitation of this work is the issue of specular glare. Glare was systematically removed during pre-processing, so that training patches did not contain any glare pixels. This was done to avoid any training biases or error that could have been introduced from glare. Moreover, since glare was removed from training, it was also removed from the testing dataset to ensure that the quantified results were unblemished by

glare artifacts to fully evaluate the classification potential of HSI. Regions that were not classified due to large amount of surface glare can be seen as grayscale in Figure 6-5.

6.5 Conclusion

In summary, this chapter presented preliminary results for this proof-of-principle experiment, which demonstrates that an HSI-based optical biopsy method using CNNs can provide multi-category diagnostic information for normal head and neck tissue and thyroid carcinomas. Future work involves acquiring and processing more patient HSI data for these types of tissue and to conduct studies of more tissue types and potentially produce results with a more universal application. Additionally, this line of research will be important in determining if the proposed technique has the potential to aid surgeons, providing a wealth of information just as histological analysis, as our preliminary results suggest.

CHAPTER 7. HEAD & NECK CANCER DETECTION IN DIGITIZED HISTOLOGY

This chapter presents the development and evaluation of a computer-aided diagnostic algorithm with the goal of aiding surgical pathologists in the detection of cancer in digitized whole-slide histological images. Pathologists guide surgeons during head and neck cancer resection operations by performing microscopic analysis of histology slides made from the excised tissue. In this study, 381 digitized, histological whole-slide images (WSI) from 156 patients with head and neck cancer were used to train, validate, and test an inception-v4 convolutional neural network. The proposed method can detect and localize primary head and neck SCC on WSI with an AUC of 0.916 for patients in the SCC testing group and 0.954 for patients in the thyroid carcinoma testing group. Moreover, the proposed method can diagnose WSI with cancer versus normal slides with an AUC of 0.944 and 0.995 for the SCC and thyroid carcinoma testing groups, respectively. For comparison, we tested the proposed, diagnostic method on a publicly-available dataset of WSI from sentinel lymph nodes with breast cancer metastases, CAMELYON 2016, to obtain patch-based cancer localization and slide-level cancer diagnoses. The experimental design yields a robust method with potential to help create a tool to increase efficiency and accuracy of pathologists detecting head and neck cancers in histological images.

7.1 Introduction

Head and neck cancer is the sixth most common cancer world-wide, and majority of cancers of the upper aerodigestive tract are squamous cell carcinoma (SCC) [2].

Approximately two-thirds of SCC patients present with either stage III or IV advanced disease [141]. Surgical cancer resection remains the primary management SCC of the head and neck, with concurrent chemotherapy or radiation therapy depending on the extent of the disease [4].

Commonly, the safe margin for surgical resection of oral squamous cell carcinoma (SCC) at sites including the surfaces of the lips, gums, mouth, plate, and anterior two-thirds of the tongue is considered 5 mm from the permanent edge of the tumor [196]. Alternative distances for resection margins have been proposed, as low as tumor clearance of 2.2 mm to be declared a "negative" margin [197]. However, closer margins, for example within 1 mm, are associated with significantly increased recurrence rates [258]. In head and neck surgical histology, there are two techniques to investigate surgical margin clearance: perpendicular sectioning and *en-face* technique. Perpendicular sectioning, also known as "bread loafing", allows the margin clearance from the edge of the resected tissue to be easily quantified, but it is resource exhaustive and limited by the number of slices, which can create false negatives. The *en-face* technique evaluates the surface area in a longitudinal fashion to determine if there is any cancer on surface of the submitted specimen [196]. Both require examining a large quantity of histological slides and considerable diagnostic time.

Previous studies have implemented computer-assisted detection methods using histological images for machine learning [259]. Colorectal epithelial and stromal tissues have been classified on histological images using support vector machines with hand-crafted features, such as color and texture [260, 261]. Additionally, convolutional neural networks (CNNs), which are a family of machine learning algorithms that learn to extract

features from training images, have also been applied to classifying epithelium and stromal tissues from colorectal and breast cancers [262]. Non-small cell lung cancers, including metastatic SCC to the lungs, have been classified in histological images using CNNs that are trained to work regions of the image, called image-patches [263]. Another method for detecting lung cancers in histological images of needle core biopsies used morphological and color features for classification with an ensemble of artificial neural networks [264]. Head and neck SCC was investigated once before, but only in cell lines xenografted into mice, and a CNN was implemented with histological images to predict hypoxia of tumor-invaded microvessels [259, 265]. Additionally, computerized methods have been developed for thyroid carcinomas to detect and classify malignant versus benign nuclei from thyroid nodules and carcinomas, including follicular and papillary thyroid carcinomas, in histological images on a cellular level with promising results [266-268]. However, most of the work involving thyroid carcinoma has been implemented on a cellular or nuclear level using hand-crafted features, such as texture or shape, and support-vector-machines are employed for nuclei classification, with many algorithms using an ensemble of classifiers [267-271].

In the field of digital pathology, whole slide imaging (WSI) refers to the acquisition of high-resolution images of stained tissue slides, which retains the ability to magnify and navigate these digital slides just as standard microscopy [272]. After reviewing nearly 2,000 patient cases, it has been concluded that WSI is non-inferior to microscopy for primary diagnosis in surgical pathology across multiple staining types, specimen types, and organ systems [273]. Computer-assisted detection algorithms have recently been implemented using CNNs for diagnosis in WSI with considerable success for identifying

metastasis in lymph nodes [274, 275]. Several state-of-the-art methods using CNNs have been applied during a grand challenge hosted at the IEEE International Symposium for Biomedical Imaging in 2016 and 2017 to detect breast cancer metastasis in WSI of sentinel lymph nodes (CAMELYON) with AUCs up to 0.99, comparable to expert pathologists performing with an AUC of 0.81 to 0.97, with and without a time constraint [274, 276, 277].

This study aims to investigate the ability of CNNs for detecting head and neck SCC and thyroid carcinomas in a novel dataset of digitized whole-slide histological images from surgical pathology. A recent literature review shows that this is the first work to investigate SCC and thyroid carcinoma detection on a WSI level in primary head and neck cancers [259], and we implement state-of-the-art classification methods in an extensive dataset collected from our institution. The major contribution of this paper focuses on the first application of deep learning for the histological detection of H&N SCC and thyroid cancers in a sufficiently large head and neck cancer dataset that is best suited for a patch-based CNN approach. The anatomical variation of the head and neck is astonishingly complex. The inclusion of multiple, most common locations of SCC yields a successful and substantial generalization for this application. Additionally, three of the major forms of thyroid carcinoma are studied, and despite extensive morphological differences, the method allows successful performance. Altogether, the dataset and applied methodology of this work demonstrate the current potential to create a tool to increase the efficiency and accuracy of surgical pathologists performing real-time SCC cancer detection on WSI for intraoperative guidance during primary head and neck cancer resection operations.

7.2 Materials and Methods

In this section, the materials for this study, including the cancer histological datasets, are described. Additionally, the methods of image processing, convolutional neural networks, and performance evaluation are detailed.

7.2.1 Head and Neck Cancer Dataset

Informed, written consent was obtained from all patients consented for our study. All experimental methods were approved by the Institutional Review Board (IRB) of Emory University under the Head and Neck Satellite Tissue Bank (HNSB, IRB00003208) protocol. In collaboration with the Otolaryngology Department and the Department of Pathology and Laboratory Medicine at Emory University Hospital Midtown, freshly excised, *ex-vivo* head and neck cancer tissue samples were obtained from previously consented patients undergoing surgical cancer resection [113, 147]. Tissue specimens collected from patients were de-identified and coded by a clinical research coordinator before being released to our laboratory for research purposes only. Three tissue samples were collected from each patient: a sample of the tumor, a normal tissue sample, and a sample at the tumor-normal interface.

For this study, we present the first application of the histological component of this dataset of 381 WSI from 156 patients, which is detailed by dataset in Table 7-1. In the upper aerodigestive tract SCC group, there were 228 tissue samples collected from 97 patients. The number of patients and tissue specimens is enumerated per anatomical origin of the SCC in Table 7-2. The only tissues that were excluded in this study were from three patients that had SCC of Waldeyer's ring. These tissues were excluded because they were

comprised of entirely lymphoid tissue, and the samples from only 3 patients of this diverse tissue type was not sufficient for inclusion in this study. The normal specimens collected were non-dysplastic and non-cancerous, which may have inflammation, atypia, or reactive epithelium.

Table 7-1. Summary of the number of patients and whole-slide images (WSI) included in this study for training, validation, and testing of the proposed method.

Dataset	Training		Validation		Testing		Total	
	Patients	WSI	Patients	WSI	Patients	WSI	Patients	WSI
Head & Neck SCC	45	91	13	32	39	105	97	228
Thyroid Carcinoma	24	48	8	23	27	82	59	153
Breast Cancer Mets.	250	250	20	20	129	129	399	399

The thyroid carcinoma group was comprised of primary papillary, medullary, and follicular thyroid carcinomas. There were 153 tissue specimens collected from 59 patients, which included 47 patients with papillary thyroid carcinoma, 5 patients medullary thyroid carcinoma, and 7 patients with follicular carcinoma. Each dataset was subdivided into separate groups for training, validation, and testing of the proposed computer-assisted cancer detection algorithm.

Fresh *ex-vivo* tissues were collected from the surgical pathology department and fixed, paraffin embedded, sectioned, stained with haemotoxylin and eosin (H&E), and digitized using whole-slide scanning at an equivalent magnification to 40x objective using a NanoZoomer (Hamamatsu Photonics), which produces a final digital slide with pixel-level resolution of 0.23 μm × 0.23 μm. A board-certified pathologist with expertise in H&N pathology outlined the cancer margins on the digital slides using Aperio ImageScope (Leica Biosystems Inc, Buffalo Grove, IL, USA).

Table 7-2. Summary of the number of patients in the SCC dataset and WSI obtained from tissue specimens per anatomical location of the head and neck. Tissue specimens refer to the ex-vivo samples used to construct the histological WSI (T: tumor, N: normal, TN: tumor-normal interface).

Location	# Patients	# T	# N	# TN
Tongue	18	9	17	17
FOM	12	7	10	13
Larynx	10	10	9	3
Mucosal Gingiva	9	7	9	4
Mandible	8	6	6	5
Maxillary Sinus	6	4	6	5
Oral Cavity	6	4	7	6
Hypopharynx	5	4	5	1
RMT	5	6	3	5
Tonsil	5	4	3	6
Supraglottis	4	2	4	3
BOT	4	0	4	4
Nasal Cavity	2	2	1	0
Other	3	2	2	3
Total	97 Patients	228 WSI		

7.2.2 Breast Cancer Lymph Node Metastases Dataset

For external validation, we implemented the proposed cancer detection algorithm on the open-source CAMELYON 2016 dataset [275, 278], in order to compare the results of our proprietary head and neck cancer dataset since currently no similar independent dataset exists. The CAMELYON 2016 dataset consists of 399 whole-slide digital images from sentinel lymph nodes (SLN) obtained from 399 patients, one SLN from each patient that underwent breast cancer surgical resection. The dataset is collected at two institutions: Radboud University Medical Center (RUMC) Netherlands and University Medical Center

Utrecht (UMCU) Netherlands [275, 278]. One slide was constructed from one SLN from each patient. Table 7-1 shows the numbers of patients and slides in each group.

The whole-slide images were digitized at each institution separately, so the different hospitals each use a different scanner. The slides that were digitally scanned at RUMC were produced at 20× objective magnification using a Pannoramic 250 Flash II digital slide scanner (3DHISTECH), which corresponds to the pixel-level resolution of $0.24\ \mu m \times 0.24\ \mu m$. The slides that were digitized at UMCU were acquired with a NanoZoomer-XR digital slide scanner at 40x objective magnification (Hamamatsu Photonics) with a pixel-level resolution of specimens of $0.23\ \mu m \times 0.23\ \mu m$ [275, 278].

7.2.3 Histological Image Processing

The histological dataset presented consists of primary tumor specimens acquired from surgical resections. Our SCC and thyroid cancer datasets do not have fine cellular-level annotations. Instead, regions were broadly marked as cancer if there were any cancer cells present, even if surrounded by normal structures, to establish which areas would require surgical removal. For this task, cell-by-cell annotations are not necessary. Clinicians require accurate regional diagnosis of cancer invaded tissues with an estimate of border clearance distance to the edge of the resected tissue. Therefore, the nature of the ground-truth for this work necessitates a patch-based deep learning approach. Moreover, a fully-convolutional network (FCN), as is widely used in the literature, would be problematic for this approach. Firstly, the tissue specimens of primary cancers collected tend to have large regions of each class. Therefore, the large majority of patches tend to be one class (all normal or all tumor), with few border patches that contain both classes. This

would create problems with loss calculation and gradient optimization for training an FCN. Lastly, as stated the ground truth is coarse, so if a FCN could be adequately trained to produce fine-level segmentations, not only are they not needed for this task, but the ground truth would call potentially correct areas as misclassifications.

A ground-truth binary mask of the cancer area is produced from each outlined histology slide. The WSIs and corresponding ground-truths were down-sampled by a factor of four using nearest neighbor interpolation. The proposed method classifies the WSI in a patch-based method using a window that slides over the entire image. Due to the unique challenges of working with digital pathology images, which can create datasets of hundreds of images that are each tens of gigabytes, it is the current state-of-the-art to perform both down-sampling and patch-based image reconstruction approaches to computationally handle this type of data [112, 274, 279-284]. Image patches ($\boldsymbol{I}$) are produced from each down-sampled H&E slide using 101×101 pixels and are labeled corresponding to the center pixel, where $\boldsymbol{I} \in \mathbb{R}^{101\times101\times3}$. Representative patches from H&N SCC are shown in Figure 7-1 showing the histological variation of normal anatomical structures and various appearances of SCC of various identifiable difficulty. The SCC and thyroid carcinoma training groups were comprised of patches only from the tumor and normal tissue WSI, and the validation and testing groups were comprised of patches from all slides. Since the lymph node dataset contained more WSI but with smaller cancer areas, the training dataset was constructed by taking up to 5000 image patches from the cancerous area of each of the 101 cancer WSI in the training dataset, and using up to 1000 image patches from each slide of the 149 normal WSI. The training group was approximately balanced between cancer and normal patches for better performance.

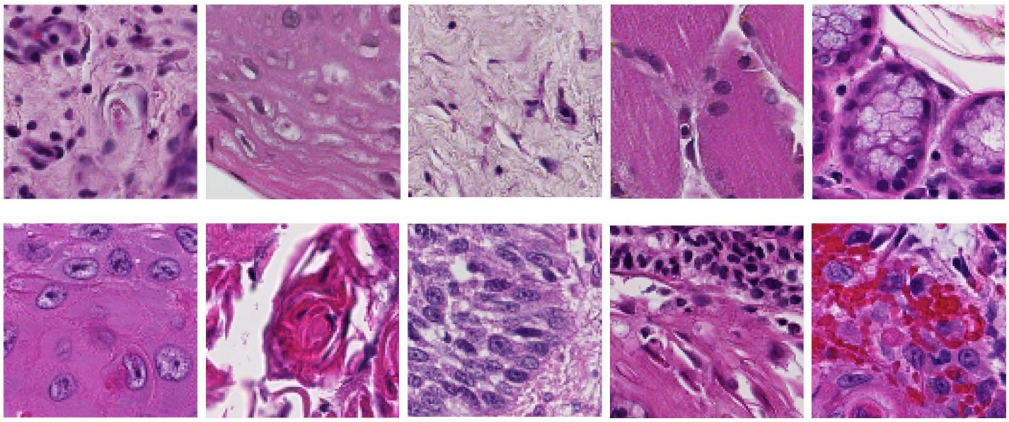

Figure 7-1. Histological images (101×101 pixel image-patches) showing anatomical diversity. Top: Patches of various normal structures, including chronic inflammation, stratified squamous epithelium, stroma, skeletal muscle, and salivary glands (from left to right). Bottom: Patches of SCC with varying histologic features: keratinizing SCC, keratinizing SCC with keratin pearls, basaloid SCC, SCC with chronic inflammation, SCC with hemorrhage (from left to right).

Histology slides have no canonical orientation, meaning the tissue will have the same diagnosis from all vantage points. Therefore, the number of image patches were augmented by 8× by applying 90-degree rotations and reflections to develop a more robust diagnostic method. Additionally, to establish a level of color-feature invariance and tolerance to differences in H&E staining between slides, the hue, saturation, brightness, and contrast of each patch were randomly manipulated to make a more rigorous training paradigm.

7.2.4 Convolutional Neural Network Implementation

The three distinct cancer datasets in this study were employed to separately train, validate, and test a 2D-CNN classifier based on the Inception V4 architecture, implemented in TensorFlow on 8 Titan-XP NVIDIA GPUs [27, 192, 217, 285]. The Inception V4 CNN

architecture was modified slightly in the early layers, which is detailed in Table 7-3, in order to accommodate the patch-size selected for this study. The CNN architecture consisted of 3 convolutional layers and 1 max-pooling layer to accommodate the patch-size used, and in total the CNN contained 141 convolutional layers and 18 pooling layers [27, 217]. Gradient optimization was performed using the Adadelta optimizer with an initial learning rate of 1.0 that was exponentially decayed by 0.95 every 3 epochs of training data [253]. The softmax cross entropy was used as the loss function. If the k^{th} training patch is denoted as $\boldsymbol{I}_k \in \mathbb{R}^{101\times101\times3}$, $k = 1, 2, \dots, K$, where K is the number of training patches in a batch, the CNN training process is to find a function $F: \mathbb{R}^{101\times101\times3} \rightarrow \mathbb{R}$ that minimize the following cost function $\mathcal{L}$:

$$\mathcal{L} = -\frac{1}{K}\sum_{k=1}^{K}[g_k^N log(p_k) + g_k^P log(1 - p_k)]$$

where

$$p_k = \frac{e^{F(\boldsymbol{I}_k)}}{\sum_{j=1}^{K} e^{F(\boldsymbol{I}_j)}}$$

and g_k^N and g_k^P are the ground truth labels for cancer-negative and cancer-positive tissue classes, respectively, corresponding to the k^{th} patch.

Table 7-3. Schematic of the proposed modified Inception V4 CNN. The input size is given in each row, and the output size is the input size of the next row. All convolutions were performed with sigmoid activation and 40% dropout.

Layer	Kernel size / Remarks	Input Size
Conv	3×3 / 'valid'	101×101×3
Conv	3×3 / 'valid'	98×98×32
Max Pool	2×2 / stride=2 'valid'	96×96×64
Conv	3×3 / stride=2 'valid'	48×48×64
4 x Inception-A Block	1×1 and 3×3 / 'same'	23×23×80
Reduction-A Block	1×1 and 3×3 / 'same'	25×25×384
7 x Inception Block	1×1, 1×7, 7×1, and 3×3 / 'same'	11×11×1024
Reduction-B Block	1×1, 1×7, 7×1, and 3×3 / 'same'	11×11×1024
3 x Inception-C Block	1×1, 1×3, 3×1, and 3×3 / 'same'	5×5×1024
Avg. Pool	5×5 / 'valid'	5×5×1536
Linear	Logits	1×1536
Softmax	Classifier	1×2

The validation groups were used to determine the optimal number of training epochs used for each of the three datasets. Each CNN was trained with a batch size of 128 image patches, and batches were converted from RGB to HSV before being passed into the CNN. Both RGB and HSV were tested in early validation experiments, and HSV without any other modification out-performed RGB results. One reason could be the separation of the image intensity from the color information in HSV color model. Additionally, one major challenge of H&E stained images is inconsistency of the stain quality. To demonstrate that color feature augmentation can solve this problem, working in HSV directly, the hue, saturation, and brightness were perturbed randomly in each channel independently. The SCC CNN was trained for 30 epochs of training data, equivalent to 295,000 steps using a batch-size of 128 patches. The random color augmentation was using the native color feature variance in the training group: hue 4%, saturation 15%, brightness 8%, and contrast 2%. The thyroid carcinoma CNN was trained

for 70 epochs of training data (equivalent to 433,000 steps). HSV and contrast perturbation was 5%, 5%, 8%, and 5% respectively. The breast cancer SLN metastasis CNN was trained for 20 epochs (equivalent to 203,400 steps). HSV and contrast were each randomly perturbed in range of -10% to 10%.

7.2.5 Image Reconstruction and Post Processing

Each of the N testing slides ($\boldsymbol{S}^{T_t}, t = 1,2, \dots, N$) were tiled into image patches of size 101×101 pixels ($\boldsymbol{I}_{m,n}^{T_t}$), produced with a stride of 50 pixels, which makes an overlap of 51 pixels.

$$\boldsymbol{I}_{m,n}^{T_t} = \{\boldsymbol{S}^{T_t}(x, y) | m \leq x \leq m + 100, n \leq y \leq n + 100\}$$

where $m = 1,51,101, \dots, M_t$, $n = 1,51,101, \dots, N_t$, and $M_t \times N_t$ is the size of the t^{th} testing slide. Therefore, the final classified whole-slide image is a cancer probability heat-map with a level of detail equal to 50×50 pixel blocks. Each image patch, $\boldsymbol{I}_{m,n}^{T_t} \in \mathbb{R}^{101 \times 101 \times 3}$, was classified in all 8 orientations with randomized HSV and contrast features and averaged to obtain a single cancer prediction value, according to the following equation.

$$\bar{p}(\boldsymbol{I}_{m,n}^{T_t}) = \frac{1}{8} \sum_{s=1}^{8} p_s(\boldsymbol{I}_{m,n}^{T_t})$$

where $p_s(\boldsymbol{I}_{m,n}^{T_t})$ is the cancer prediction for the s^{th} orientation of $\boldsymbol{I}_{m,n}^{T_t}$ and $\bar{p}(\boldsymbol{I}_{m,n}^{T_t}) \in [0,1]$ is the average cancer prediction of the patch. Additionally, the results of overlapping image patches were averaged in the overlapping area, as follows.

$$\bar{q}(\boldsymbol{H}_{i,j}) = \frac{1}{4}\left(\bar{p}(\boldsymbol{I}_{i,j}^{T_t}) + \bar{p}(\boldsymbol{I}_{i+49,j}^{T_t}) + \bar{p}(\boldsymbol{I}_{i,j+49}^{T_t}) + \bar{p}(\boldsymbol{I}_{i+49,j+49}^{T_t})\right)$$

where $\bar{q}$ is the final probability of the final resolution block size ($\boldsymbol{H}_{i,j} \in \mathbb{R}^{50\times 50}$) of the heat-map. The benefit of this post-processing method was to increase the resolution of the heat-map from 101-pixel image patches to 50-pixel image patches. Moreover, the image patches that constituted the free edge of the tissue were averaged less than four times because they did not have the complete number of neighboring patches. This image reconstruction and post-processing method was determined to increase accuracy by about 2% in early validation experiments.

To investigate the ability of the CNN to detect cancer on histological images, we implemented the gradient class-activated map (grad-CAM) method to visualize gradients activated by each class for the example input image patches [218]. We traced the gradients from the last convolutional layer before the inception modules to the logits layer to separately visualize cancer and normal components. This technique produces a weighted combination of the convolutional filters and gradients as the CNN is activated by a specific input image for each class.

7.2.6 Performance Evaluation

The reference standard cancer margin was annotated by hand for all digital slides employed in this study. For the head and neck cancer database, a board-certified pathologist with expertise in H&N pathology outlined the cancer margins on the digital slides. For the breast cancer metastasis database, an experienced lab technician and a clinical Ph.D.

student outlined the cancer margins, which were then confirmed by one of two board-certified pathologists with expertise in breast cancer [278].

During training, the performance of the validation group was calculated and monitored. The optimal operating threshold was calculated from the validation group for generalizable results, and it was used for generating performance evaluation metrics for the testing group. To reduce bias in the experiment, the fully-independent testing group was only classified a single time at the end of the experiment, after all the network optimizing had been determined using the validation set. To test the ability to diagnose and localize cancer on WSI, we used AUC, F1 score, accuracy, sensitivity, and specificity to evaluate cancer detection on a patch-based level. Confidence intervals were calculated using a boot-strapping algorithm. Additionally, the ability of the proposed algorithm to diagnose slides with cancer from normal slides was investigated. This slide-level AUC was calculated by assigning the value of the image patch with the maximum cancer probability to the entire WSI.

7.3 Results

7.3.1 Digital Whole-slide Image Cancer Detection

Head and neck primary SCC was detected on digitized WSI with an AUC of 0.916 and 85% accuracy for patients in the testing group. The ideal threshold for distinguishing SCC from normal tissue was SCC probability of greater than 62%. Thyroid carcinoma was detected on digitized WSI with an AUC of 0.954 and 89% accuracy for patients in the testing group. The ideal threshold for distinguishing thyroid carcinoma from normal thyroid tissue was cancer probability of greater than 50%. Breast cancer lymph node

metastasis was detected on digitized WSI with an AUC of 0.967 and 93% accuracy for patients in the testing group. The ideal threshold for identifying metastasis in SLNs was cancer probability of greater than 28%. Reported in Table 7-4 are the AUC for the validation groups and the AUC, accuracy, sensitivity, and specificity of the testing groups.

Table 7-4. Cancer detection results, obtained from ROC curves using all histological images' patch-level statistics. Reported are the AUC for the validation group and the AUC, F1 score, accuracy, sensitivity, and specificity of the testing group with 95% confidence intervals for all values. Also shown in the right-most column is the ability of the proposed method to distinguish slides that contain cancer from slides that are all normal.

Group	Validation AUC	Test AUC	F1 Score	Accuracy	Sensitivity	Specificity	Slide Level AUC
SCC	0.913 (0.90,0.93)	0.916 (0.90, 0.93)	84.8±1.5%	84.8±1.6%	84.7±2.2%	85.0±2.2%	0.944 (0.91, 0.97)
Thyroid	0.927 (0.92, 0.94)	0.954 (0.94, 0.97)	89.4±1.3%	89.4±1.3%	89.6±1.8%	89.1±1.9%	0.995 (0.99, 1.00)
Lymph Node	0.986 (0.96, 0.99)	0.967 (0.96, 0.98)	91.8±1.3%	93.4±1.2%	90.1±1.8%	93.6±1.6%	0.901 (0.86, 0.94)

Receiver operator characteristic (ROC) curves for slide-level and patch-level cancer detection in the testing groups from all three datasets are shown in Figure 7-2. Patch-level ROC curves are generated using all histological images' patch-level data for cancer localization, and slide-level ROC curves demonstrate WSI diagnosis. Additionally, two representative WSI from each of the three testing groups and their corresponding predicted heat-maps are shown in Figure 7-3. Several regions of interest (ROI) are detailed in Figure 7-4 to identify the strengths and weaknesses of the proposed method in the detection of SCC. The ideal threshold for whole-slide level detection of SCC was above 95% probability, so the regions detailed as true negatives in Figure 7-4 fall below this threshold. Additionally, the grad-CAM technique was used to visualize the contributing normal and cancerous components of a few example input images that were corrected classified with

high probability (Figure 7-5). This approach reveals that a contribution of the cancer prediction is made by nuclear features.

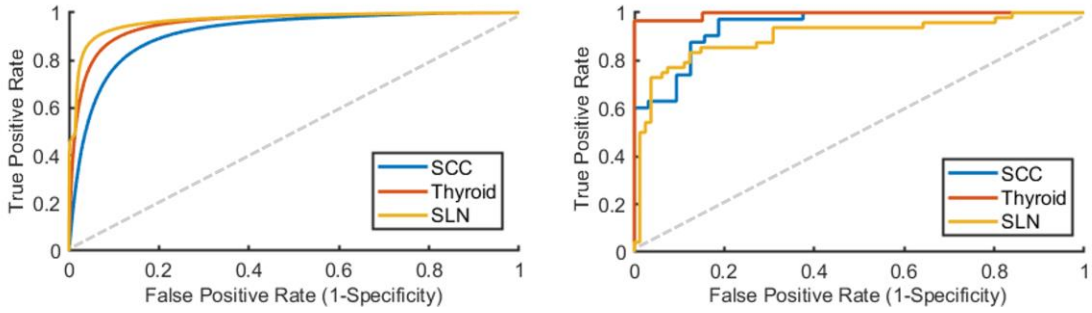

Figure 7-2. The ROC curves are shown for patch-level cancer detection (left) and slide-level cancer diagnosis (right) in the testing groups from all three datasets. The dotted gray line corresponds to random guess.

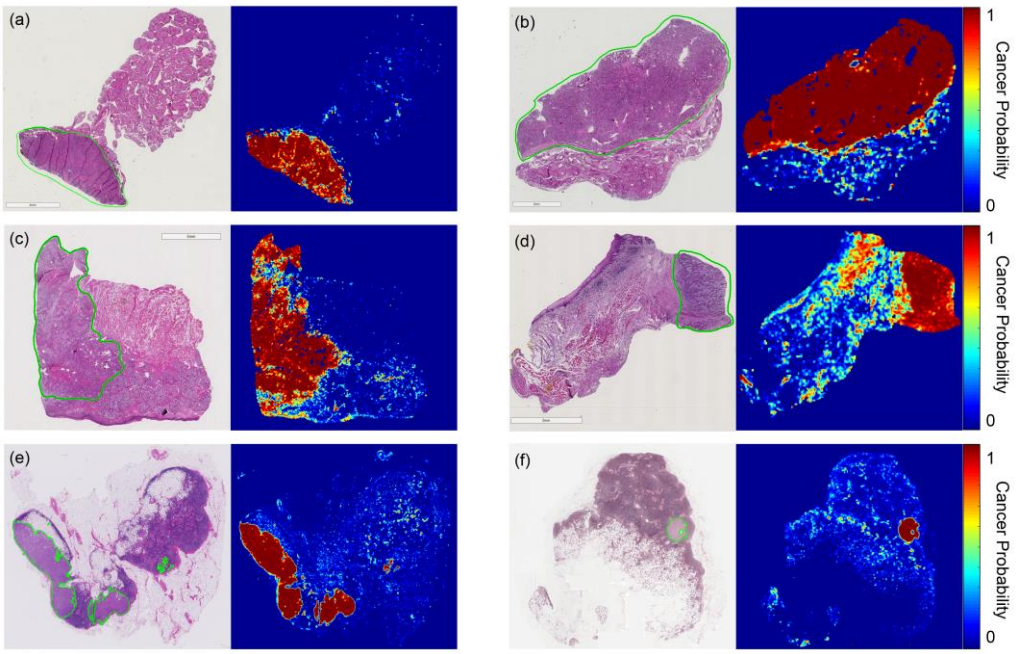

Figure 7-3. Representative whole-slide classification results. (a, b) Papillary thyroid carcinoma WSI from two patients. (c, d) SCC WSI from patients with tongue and retromolar trigone SCC. (e, f) Breast cancer metastasis to lymph node WSI from two patients. The cancer area is outlined in green on the H&E images, and the heat maps are shown of the cancer probability.

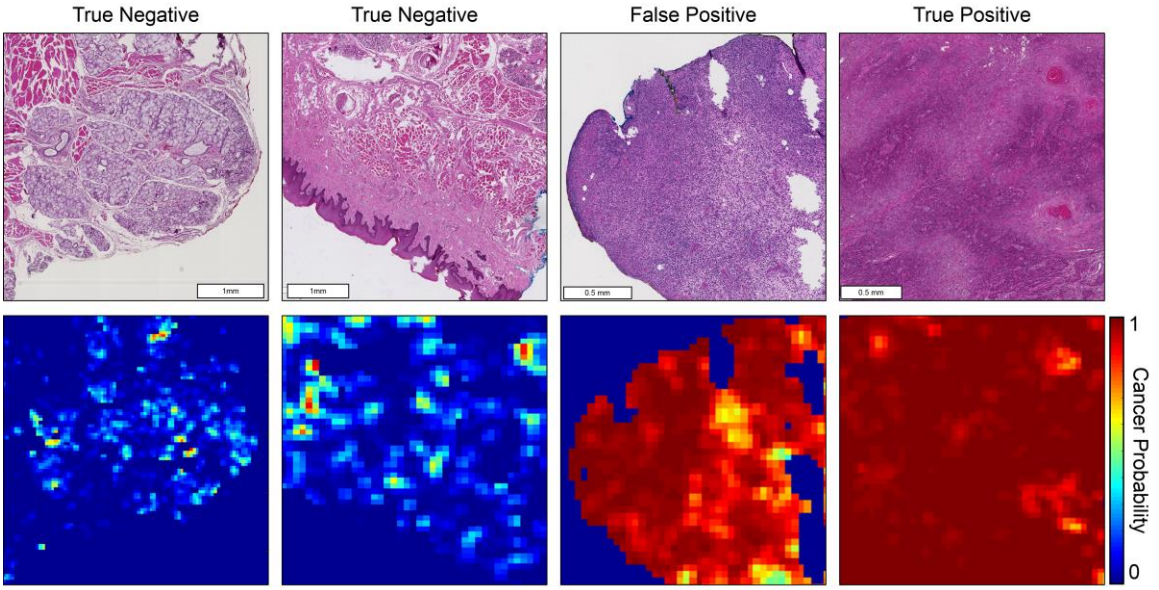

Figure 7-4. Heat maps representing cancer probability of several regions of interest. From left to right, the CNN correctly identifies salivary gland and muscular components as having a low probability of SCC; stratified squamous epithelium correctly shown as a true negative; a false positive area representing inflammatory infiltration near the SCC border (not shown); correctly classified true positive SCC classified with a high probability of SCC.

The ability of the proposed method to diagnose the entire WSI that contain any cancer was also investigated. WSIs with SCC were diagnosed with an AUC of 0.944. Thyroid carcinoma WSIs were diagnosed with an AUC of 0.995. WSI of SLN with breast cancer metastases were diagnosed with an AUC of 0.901.

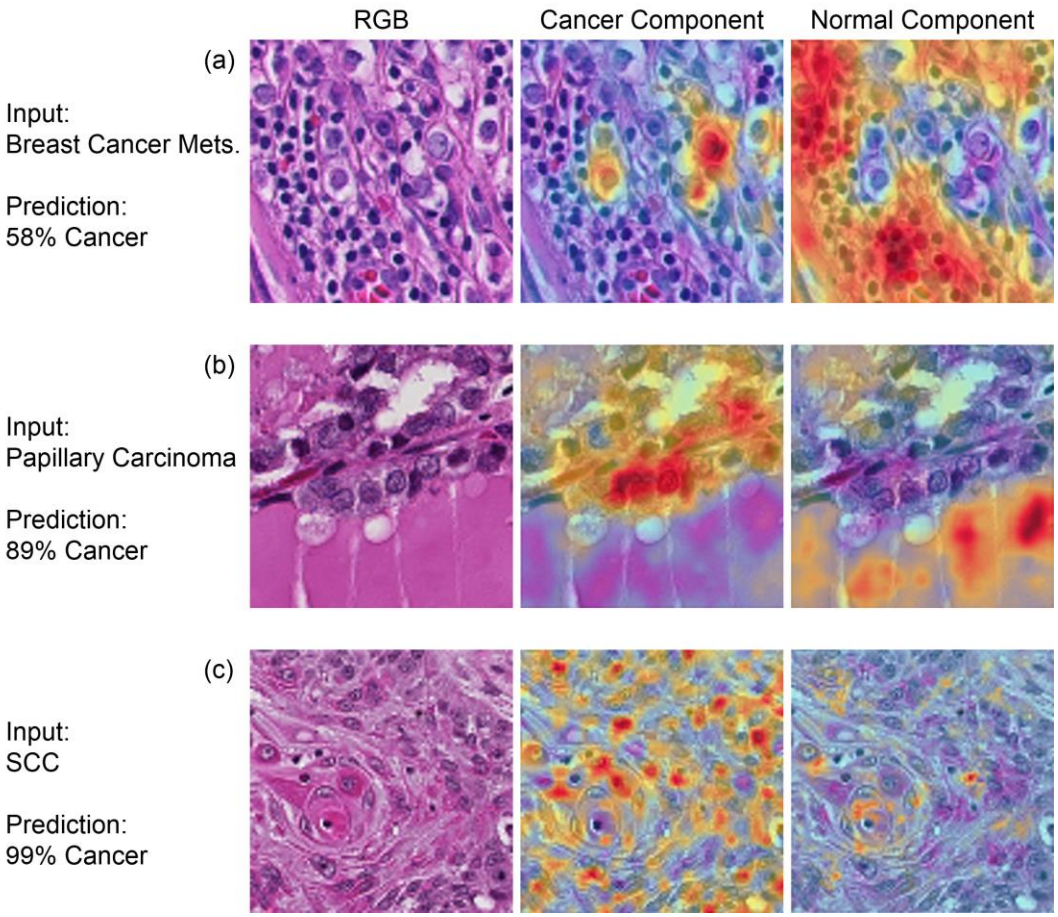

Figure 7-5. Representative, correctly-classified cancer image-patches from each dataset, visualized with cancer and normal components using the grad-CAM technique. The gradients were traced from the last convolutional layer before the inception modules to the cancer and normal logits separately to visualize contributions from cancer and normal features leading to a correct cancer prediction. (a) Breast cancer SLN metastasis image patch correctly classified as cancer with 58% probability. (b) Papillary thyroid carcinoma image patch correctly classified as cancer with 89% probability. (c) SCC image patch correctly classified as cancer with 99% probability.

7.4 Discussion

In this work, we present a new and extensive histological dataset of primary head and neck cancer and implement a state of the art Inception V4 CNN architecture for cancer detection and WSI diagnosis. The results are generalizable because of the division of patients across training, validation, and testing. To the best of our knowledge, this is the

first work to investigate SCC detection in digitized whole-slide histological images from primary head and neck cancers.

The digitized, whole-slide histological images were saved as TIF files with resolution equivalent to 40x microscopic objective. After 4x down-sampling, the image patches correspond to 10x objective equivalence. Different down-sampling factors and patch-sizes were explored, but this method yielded the best validation group results, so it was used for testing. Similarly, pathologists detecting SCC in histology slides use a variety of objectives, not exclusively 40x, which may be too zoomed-in to determine if the region is cancerous or benign. We see this issue in our dataset as well. It is not only possible, but likely that in some slides labeled as 'tumor only', there may be some areas inside the tissue, or in between tumor nests, that is entirely normal. Therefore, it is understandable that classification using 4x down-sampled images obtains the highest accuracy. Other CNN architectures were explored in early experiments using the validation set only, and various patch-sizes were experimented with, but ultimately the Inception V4 CNN architecture with a patch-size of 101×101×3 in HSV color space, yielded the most promising validation results.

Additionally, the regions of interest that are presented show true negative, false positive, and true positive regions that vary from 1 to 3 mm in size. These results demonstrate the proposed method is able to distinguish normal anatomical structures like epithelium and salivary gland from SCC with high probability. Also, the most common false positive observed in the classified result is tissue areas that contain dense inflammation. This result is most likely a by-product of the training paradigm. As SCC develops, there is an accompanying immune response that leads to massive inflammatory

infiltration into the tissue [286]. Therefore, the proposed algorithm learned the association between SCC and inflammation.

To our knowledge, there are no other studies that attempt to detect or diagnose H&N SCC or thyroid carcinoma on WSI, and we used a proprietary dataset collected from patients at our institution. Therefore, we wanted to test the proposed, diagnostic algorithm on a similar, open-source dataset for comparison. Our slide-level results would have placed 3rd in the original CAMELYON 2016 [275, 278].

The grad-CAM technique was used to visualize what components of the input image are determined as useful features with a significant contribution to the cancer prediction from the CNN, as shown in Figure 7-5. This reveals that the decision is made by looking at the nuclei, just like a pathologist detects cancer. The proposed method does not segment all cancerous nuclei in the image patch, but it identifies a few cancerous nuclei with a high probability of being cancer and uses this information for making the decision. We did not train the proposed algorithm specifically with this in mind. Rather, this phenomenon was learned naturally by the training paradigm. The trained CNN model also has a level of stain invariance.

7.5 Limitations

One limitation of the presented approach results from the application of the down-sampled resolution and the patch-based Inception V4 CNN implementation. After down-sampling, each pixel represents approximately 0.91 microns, which produces a patch size that spans about 92 microns in each x-y dimension for the patch size of 101×101 that was implemented in this approach. The typical diameter of an SCC single-cell nucleus in our

dataset was about 12 microns, which agrees with values reported in the literature of about 13 +/- 2 microns [287]. Therefore, the theoretical limit of the smallest carcinoma that could be detected would be a nest of SCC cells with an approximate diameter of 92 microns. This value corresponds with an SCC nest on the order of tens of cells of SCC, depending on the cytoplasmic overlap in the arrangement of the SCC nest.

Another limitation of this approach was that the algorithm suffered from whole-slide scanning artifacts, such as out-of-focus regions and including errors from slide processing, such as tissue folding and tearing. This was discovered after the completion of the experiment, and the effect was substantial, accounting for the reason in misclassification of the lowest performing WSI in the SCC testing dataset, which is shown in Figure 7-6. As can been in Figure 7-6, the left side of the WSI is classified correctly as a true positive SCC region, but the out-of-focus regions result in misclassification a similar ROI (shown in the cut-out boxes on the right) to be classified as false negative incorrect result. These misclassifications were retained in the testing dataset to not manipulate or bias the results, but in future work, slide scanning artifact detection should be additionally performed to determine which slides cannot be classified because of limited quality.

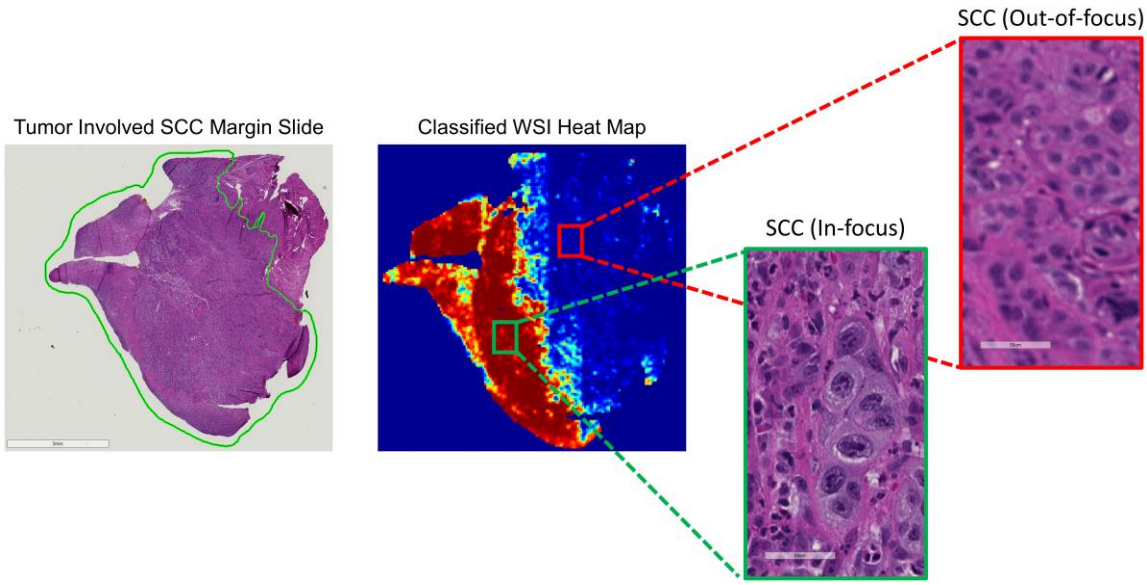

Figure 7-6. Representative false negative resulting from slide scanning artifacts. This accounts for the misclassification result of the lowest performing WSI in the SCC test group. The left side of the WSI was a correct true positive classification (green box), but the out of focus artifact results in false negative misclassification (red box).

Digital pathology with WSI allows pathologists to view high-resolution histological images, just as standard microscopy, and it was concluded that digital pathology is non-inferior to microscopy for primary diagnosis in surgical pathology cases across multiple institutions, staining types, and organ systems [272, 273]. Therefore, we believe the robust experimental procedure of the proposed method, designed to eliminate bias, has demonstrated potential benefit in a modern, digitized clinical setting. However, primary diagnosis of surgical specimens for intraoperative guidance is performed on frozen-sections rather than formalin-fixed, paraffin embedded tissues, as were investigated in this study. Additionally, frozen-sections are typically lower quality than those created from fixed, embedded specimens because they suffer from many different artifacts and depend heavily on the skill of the operator. Therefore, we believe the presented work demonstrates potential for clinical benefit, but more investigation needs to be performed.

Moreover, the generalization of the results beyond head and neck cancers to breast cancer metastasis in sentinel lymph nodes suggests this method is not limited to any organ system and could be adapted to serve multiple purposes if implemented in a more clinical setting.

7.6 Conclusion

In summary, this chapter details the first application of deep learning for the histological detection of H&N SCC and thyroid cancers. The proposed method can detect and localize primary head and neck SCC on whole-slide, digitized histological images with an AUC of 0.916 for patients in the SCC testing group and 0.954 for patients in the thyroid carcinoma testing group. Moreover, the proposed method can discriminate WSI with cancer versus normal slides with an AUC of 0.944 and 0.995 for the SCC and thyroid carcinoma testing groups, respectively. The SCC detection method is performed across all anatomical locations, which indicates the algorithm is not limited to one location of the head and neck anatomy. For thyroid cancers, three major thyroid carcinomas are studied together, which additionally demonstrates the generalizability of the method. For external validation, we tested the proposed method on a publicly-available dataset, CAMELYON 2016, and obtained good results. The agreement between validation and testing demonstrate that the technique is generalizable due to the robustness of the training paradigm and the careful experimental design to reduce bias. Together, the novel application to our dataset and promising results of this work demonstrate potential that such methods as the one proposed could help create a tool to increase efficiency and accuracy of pathologists performing head and neck cancer detection on histological slides for intraoperative guidance during head and neck cancer resection operations.

CHAPTER 8. CONCLUSION

8.1 Summary and Key Findings

8.1.1 Ex-vivo HSI Patient Database

A dataset of 204 patients with head and neck cancer was acquired (Table 8-1). Optical imaging was performed with HSI, autofluorescence, 2-NBDG dye, and proflavin dye on all the 204 patients collected for this thesis work. Digitized histological slides were made from all tissue specimens and annotated by a pathologist. The data were sorted into groups according to the patients' medical and histological information. All imaging data were calibrated, processed, and registered to the ground truth with the method described in Chapter 3. Thus, this processed dataset was prepared for analysis within this thesis work and can be used for future studies in HSI using this same dataset by other researchers.

Table 8-1. Summary of patient dataset acquired for this thesis.

HNSCC *(Chapter 4)*		***Thyroid*** *(Chapter 5)*	
SCC, Conventional	85	Thyroid, Papillary Carc.	54
SCC, HPV+ (p16+)	13	Thyroid, Follicular	13
SCC, Basaloid	1	Thyroid, Medullary Carc.	5
SCC, Neuroendocrine	1	Thyroid, Insular Carc.	1
SCC, Spindle Cell	1	Thyroid, Poorly Diff. Carc.	3
Adenosquamous Carcinoma	1	Thyroid, Benign Goiter	6
Total	**102**	**Total**	**82**
Other *(Not used)*		***Salivary Glands*** *(Chapter 5)*	
Spindle Cell Carcinoma	1	Pleomorphic Adenoma	2
Osteosarcoma	1	Mucoepidermoid Carcinoma	1
Reactive Tonsil	1	Salivary Duct Carcinoma	1
Lymph Nodes	1	Polymorphous LG Adenoma	1
Lung Adenocarcinoma	1	Adenoid Cystic Carcinoma	1
Jugular Vein SCC Metastasis	1	**Total**	**6**
Incomplete Information	8		
Total	**14**		

8.1.2 Cancer Detection with HSI

This thesis work focused on the use of deep learning, particularly deep CNNs for cancer detection using HSI input. As discussed in Chapter 3, the rationale of the selection was that CNNs can tolerate variance in the training data, learn generalizable features, can be deployed in seconds after the model is fully-trained, and have demonstrated superior performance to traditional, regression-based machine learning methods since CNNs utilize non-linear activations. In the ML methods compared in this thesis, both spectral only and spectral-spatial methods that utilized traditional regression-based ML were tested, and it was observed that the CNN could outperform them, particularly when using a large training group. However, the experiments comparing different ML algorithms were intentionally limited in number and performed on a subset of the patients to reduce the substantial bias that could arise from "over-testing" with our only large dataset.

In Chapter 4, we reported on the investigation of deep CNNs for SCC detection and the performance of label-free, reflectance-based HSI and autofluorescence imaging in 293 tissue specimens from 102 patients, comparing with fluorescent dyes. The inter-patient experiments showed that reflectance-based HSI and autofluorescence imaging outperformed proflavin dye and standard RGB images ($p<0.05$) for conventional SCC detection. Up to 2 mm near the cancer margin, in specimens from the larynx, oropharynx, and nasal cavity, HSI detected conventional SCC with 0.85–0.95 AUC score. In the oral cavity, hypopharynx, and maxillary sinus, HSI detected conventional SCC with around 0.80 AUC score. The tongue specimens were the worst performing location for SCC detection using HSI, around 0.65 AUC score, which may be a result of myoglobin absorption signals interfering with the hemoglobin absorption signal that the CNN uses for

prediction. These results confirmed the hypothesis that HSI would outperform other imaging modalities for conventional type SCC. In HPV+ SCC, autofluorescence imaging detected in tonsillar tissue with 0.91 AUC score, outperforming other modalities. This study demonstrates that label-free, reflectance-based HSI and autofluorescence imaging methods can accurately detect the cancer margin in ex-vivo specimens within minutes, which could aid surgeons and reduce inadequate surgical margins during SCC resections.

In Chapter 5, we reported on the use of CNNs for thyroid and salivary gland tumor detection with HSI and autofluorescence imaging in 216 tissue specimens from 82 patients, comparing with fluorescent dyes. For thyroid tumor detection, with average and median AUC scores of 0.86 and 0.95, HSI outperformed autofluorescence, 2-NBDG, and proflavin dyes, thus confirming the initial hypothesis. However, we observed an unexpected result. Additionally, HSI was used to synthesize three-band RGB multiplex images to represent the human-eye response and Gaussian RGBs. The average and median AUC scores of 0.90 and 0.96 results for human-eye HSI-synthesized RGB images significantly outperformed HSI ($p<0.05$). In salivary gland tumors, HSI had the best performance for parotid gland tumors with 0.92 AUC score, and autofluorescence had the best performance for other salivary glands (0.80 AUC score). This study demonstrates that HSI could aid surgeons and pathologists in detecting tumors of the thyroid and salivary glands.

A major contribution of this work was to provide a solution to the specular glare problem, shown in Chapters 4 and 5. Previously, one major limitation of HS acquisitions of wet tissue surfaces was creating specular glare from perfect reflections, which contain no useful information for diagnosis and confound ML classifiers. When using a very large number of tissues for training, the CNN methods demonstrated in both SCC and thyroid

cancers to be invariant or agnostic to specular glare with HSI, compared to RGB imaging methods, which was very susceptible to glare-induced misclassifications. Originally, specular glare was believed to be a large confounding factor with tissue classification, even in deep learning, but it was discovered that training a very deep CNN (hundreds of layers) with a large number of HSI tissue specimens (here several hundred) could solve the glare issue.

8.1.3 Tissue Sub-component Classification using HSI

In Chapter 6, we reported sub-component classification of normal aerodigestive tract tissues into skeletal muscle, salivary glands, and stratified squamous epithelium. Additionally, the ability of HSI to discriminant benign hyperplasia / goiter from two kinds of thyroid carcinoma was investigated. This chapter presented proof-of-principle results with limited patient sample size experiments to study the ability HSI to provide multi-category diagnostic information for normal head and neck tissue and thyroid carcinomas.

8.1.4 Cancer Detection on Histology Slides

In Chapter 7, we reported on the use of deep learning algorithms for cancer detection in RGB digitized whole-slide histology imaging using the histological specimens acquired in our database. The rationale for this investigation is that during operations, pathologists guide surgeons by detecting cancer in histology slides made from the excised tissue. We reported on the results of automated cancer detection in 381 digitized, histological whole-slide images (WSI) from 156 patients with head and neck cancer, using CNNs. The proposed method demonstrated ability to detect and localize head and neck SCC in WSI with an AUC score of 0.92 and with 0.95 AUC score for thyroid carcinoma.

Moreover, we reported that simplifying the task to determine if the slide has any cancer resulted in AUC scores of 0.94 and 0.99 for the SCC and thyroid carcinoma cohorts, respectively. The experimental design yields a robust method with potential to help create a tool to increase efficiency and accuracy of pathologists detecting head and neck cancers in histological images.

8.2 Limitations

One of the major limitations that should be discussed first is that the tissue specimens collected for this study were excised, ex-vivo specimens. This imposes restrictions for two main reasons: (1) the imaging acquisition settings are ideal, and (2) the ability of ex-vivo cancer detection is not necessarily equivalent to in-vivo detection. The HSI system used in this work was intended for small animal studies and ex-vivo tissues samples. Tissues were imaged in a black box with ideal illumination conditions, focusing, and no motion artifacts. Acquiring in-vivo HSI creates all of these issues commonly [168]. Additionally, in Chapter 4, we demonstrated that the oxy-hemoglobin signal is the most salient spectral component for the CNN to extract strong gradients that enable correct cancer prediction. In ex-vivo tissues, all hemoglobin will be oxygenated, which creates a distinct absorption profile from deoxygenated hemoglobin [288], which may be found more abundantly in in-vivo acquisitions.

The metabolism of cancer has a well-studied change, called the Warburg effect, which states that cancer tends to prefer metabolism with glycolysis instead of the normal, efficient oxidative phosphorylation pathway [289]. The relative tissue compositions of key

molecules, such as pyruvate (the final product of glycolysis), oxidized and reduced FAD and NAD, with spectral signatures in the range of the VNIR acquired for this study, may yield differences in-vivo compared to ex-vivo. This change could lead to in-vivo observations that could confound or enhance the ability of HSI for cancer detection. As discussed, our ex-vivo results found oxy-hemoglobin to be the most significant feature for deep learning-based predictions. In the literature on diffuse spectroscopy of breast cancer using in-vivo acquisitions, it was demonstrated that both oxy- and deoxy-hemoglobin increase significantly in cancer compared to normal tissue, but the deoxy-hemoglobin signal was more predictive [290]. Therefore, by extension, not only will the spectral signature likely change moving from ex-vivo to in-vivo, but the most significant spectral features may also change.

The ex-vivo HSI acquisition system used for this study employed a LCTF for separating spectral bands, which may have limited the performance of this study. The resolution between spectral bands was only 5 nm, which produced a coarse spectral signature, but allowed for very high spatial resolution. Currently, systems, such as the snap-scan system, are able to capture very high-quality spectral signatures, with spectral resolution around 1-2 nm, and spatial resolutions that are beyond or equivalent the 1040 by 1392 pixels used for this thesis work. These evolving technologies allow for fast acquisitions that will hopefully limit motion artifacts. Additionally, these systems are easy to focus and more compact, which are two key components necessary for clinical translation.

The deep learning-based computer-aided diagnostic algorithm for histological images presented in Chapter 7 demonstrated successful performance in head and neck

SCC, thyroid carcinomas, and breast cancer metastasis to lymph nodes. Cancer was detected in whole-slide images with pixel-level resolution, which suggests that there is clinical value is exploring these methods further. However, there were several notable limitations. Firstly, inspecting the theoretical limit of the proposed method, cancer was detectable down to micro-carcinomas with diameter of 92 microns, which may not be sufficient for all use cases. Second, the algorithm performance suffered significantly at instances of slide scanning artifacts, such as out-of-focus regions, and errors from slide processing, such as tissue folding and tearing.

The deep learning methods used for cancer detection with HSI for this thesis were tested on our dataset, which contained many samples from many sites, but was obtained from a single institution with a single HS acquisition system. The methods presented here for both HSI and digital histology must be rigorously investigated on large patient populations from multiple institutions and using different HS cameras and slide scanners. Deep learning systems are ideal for HSI-based detection because they can be rapidly deployed once fully-trained. Therefore, additionally, a streamlined user interface should be designed for straight-forward HS acquisitions, deployment, and prediction maps that can be easily interpreted by clinicians.

8.3 Future Work

The future research in the field of medical HSI should be directed in the following areas: more in-vivo patient acquisitions, fine tuning machine learning diagnostic programs, developing better HS acquisition systems, and user-friendly deployment systems.

8.3.1 In-vivo HSI Acquisition

This thesis has both reviewed current works in the literature and presented the use of HSI and optical imaging for a comprehensive and large 204 patient ex-vivo tissue specimen dataset. Future work of this thesis work on head and neck cancer involves the extension from ex-vivo imaging with tissue specimens to in-vivo intraoperative acquisitions of human patients. There are challenges, both known and unknown, in moving from the former to the latter. Once the new in-vivo data is acquired, it should be investigated if an algorithm trained on ex-vivo data, such as presented in Chapter 4 and 5, can be generalizable enough to apply to in-vivo data for testing. Additionally, it should be investigated if the ex-vivo data can serve as pre-training or transfer learning for deep learning methods to initiate weights, which would allow fine-tuning of the models with a lower learning rate, which can be achieved using a limited number of patient HS images. Therefore, even with a smaller number of in-vivo acquisitions, the work presented in this thesis using ex-vivo HSI from 204 patients will be utilized for clinical and translational purposes.

Alternatively, if there is a delay with in-vivo acquisitions during routine clinical surgical service, it would be possible to obtain human cadaveric in-situ acquisitions to establish geometry and lighting challenges and test transfer learning from ex-vivo deep learning models. More acquisitions are required to determine the clinical viability of HSI for surgical guidance, as ex-vivo alone does not provide enough information since the acquisition protocols are in ideal circumstances. Another avenue of exploration is to determine if a single predictive model for an entire organ system, such as all sub-locations head and neck in Chapter 4 or all thyroid tumor types in Chapter 5, can be sustainable and

reliable. If not, it would require investigation into subcategorizing and re-training the cancer predictive models to obtain, for example, one model for oral cancers, one model for laryngeal cancers and so on. However, since deep learning allows meaningful feature extraction and learning from large and varied datasets, we hypothesize that a broad organ system approach as was implemented will be proven to be acceptable.

8.3.2 Emerging HSI Technology

While the use of HSI for clinical application is still in its adolescence, the development of new and emerging HSI will undoubtedly help accelerate the translation. Firstly, as described above, the HS cameras are becoming higher quality, both in spatial and spectral resolution, with faster acquisitions. However, currently a specific HS camera is needed for a specific window of the electromagnetic spectrum, such as visible, VNIR, NIR, or SWIR. The development of HS cameras that acquire shorter wavelengths than visible light (below 400 nm) needs to be established to allow for spectroscopy of signals like NAD and keratin, which are less than 350 nm. Additionally, extending the signal of HS cameras beyond 1000 nm for these same cameras would greatly help clinical utility. The depth penetration of NIR light is up to tens of millimeters, which is significantly deeper than shorter UV and visible light, which ranges from one-tenth to only a few millimeters.

In a clinical scenario, it is challenging and unrealistic to expect multiple acquisitions on multiple HS cameras in the operating room, so this process needs to be streamlined into one HS camera. Additionally, the computer aided diagnostic programs need to be packaged in operator friendly interfaces. The current acquisition programs and computer diagnostic programs for surgical guidance are cumbersome and finicky without

expert knowledge, which would stall adoption. The hurdles to both HS acquisition and deployment need to be addressed. If this happens, the wealth of HS data that can be captured could lead to better performance and accuracy across more types of cancer in other organ systems as well.

8.3.3 Computer-Aided Histology Detection

Current practice for pathological consultation during head and neck surgeries involves processing specimens using frozen-sections and standard microscopy. A recent study found that digital pathology, navigating digital images of slides in a fashion equivalent to standard microscopy, was non-inferior for primary diagnosis [273]. Therefore, if pathologists and surgeons adopt digital histology for making diagnoses, the deep learning models developed in this thesis in Chapter 7 could aid in the clinical workflow with quantitative cancer detection. On the other hand, even if pathologists do not adopt digitized slides for reading, for example because it adds time, microscopes are being developed that scan simultaneously and seamlessly without interrupting standard pathology workflows to instantly detect cancer in the microscope field of view and apply an augmented reality overlay for cancer visualization [291].

The wealth of histological data available to clinics can be utilized for training models with reduced expert annotations. Specifically, training deep learning approaches can be automated even for training, without requiring human pathologist annotations, simply needing one label per slide [292]. Reducing the work needed by expert pathologists could make it possible to create a training dataset with tens of thousands of slides from multiple institutions. Therefore, to make the translational leap from proof-of-concept to

improving surgical pathologists' clinical workflow, thousands of slides should be made available for training and more research should be conducted to these new digital microscopy methods.

8.3.4 Summary of Future Research Directions

In conclusion, several future research steps are outlined, which are necessary for the clinical translation of hyperspectral imaging: testing of new and emerging HS cameras for better acquisition systems, collecting cadaveric or in-vivo HS acquisitions, exploring how this affects CNN-based deep learning algorithms, rigorously investigating on large patient populations from multiple institutions, and streamlining the deployment of these systems clinically. Additionally, future steps needed for translation of deep learning for assisting surgical pathologists are outlined. However, the methods developed and presented in this thesis are not exclusive to HSI or histological imaging, and this work can be of value to any aspiring optical imaging modality for translation surgical guidance. Indeed, it is possible that future and emerging optical imaging modalities currently under investigation by the scientific community could be combined with the methods developed in this thesis to offer equivalent or superior cancer detections abilities for guiding surgeons in the operating room.

APPENDIX A. HISTOLOGICAL GRADING OF SCC USING HYPERSPECTRAL IMAGING

In this appendix section, we explore the predictive ability of HSI beyond cancer detection to investigate the use HSI for predicting histological SCC grade. Histological grading of SCC of the head and neck has been correlated with disease prognosis and patient survival. SCC is graded from well, moderately, to poorly differentiated (grades 1, 2, and 3, respectively). The objective of this project is to determine the ability of label-free, non-contact HSI for predicting histological grade of SCC in gross-level ex-vivo tissue specimens. Using tissue specimens from patients with known histological grade from the head and neck SCC dataset (N=56 patients, N=73 tissue specimens), we trained a CNN to predict the SCC grade. Validation and testing was performed on 22 tissues specimens from 15 fully-independent patients to determine performance. Overall accuracy for predicted grade of tissue specimens was 67% and 60%, which is significantly higher than randomly guessing SCC grade (33%). Agreement, measured with Cohen's kappa statistic, was 0.50 and 0.44, indicating a fair to moderate agreement. The inter-reader reproducibility of SCC grading is about 0.60 in the literature. The preliminary results of this work show that HSI may hold potential for automatic grading of head and neck SCC of gross-level tissue specimens. This work warrants future investigation into the comparison of the automated methods as performed with digitized histology of the same specimens to determine the agreement of grade predictive algorithms with different imaging modalities.

A.1 Introduction

Label-free HSI is a non-contact, non-ionizing, optical imaging modality that acquires a sequence of two-dimensional images over discrete wavelengths (2-5 mm) in a broadband electromagnetic spectrum, typically from 400 to 900 nm [14, 211]. Originally implemented in biomedical imaging for calculating vascularization and hemoglobin saturation, HSI has been recently applied to cancer detection, in multiple areas, including brain cancer and head and neck cancer [113, 166, 169, 179].

Head and neck cancer is the sixth most common cancer worldwide, and surgery remains the well-established standard treatment. Recurrence rates for cancer after surgical resection are largely dependent on negative, cancer-free surgical margins, along with several factors that cannot be controlled, such as histological cancer grade and staging of the cancers [293, 294]. Nearly two-thirds of patients with head and neck SCC present with advanced, late stage disease.

The histological grade of head and neck SCC has been studied as an independent prognostic factor in Cox's multivariate survival analysis ($p < 0.01$) [293]. Several studies have calculated the inter-pathologist agreement for determining histological SCC grading, with Cohen kappa values of agreement ranging from 0.60 to 0.64 [293, 295, 296]. Therefore, current practice, which involves performing histological preparation of *ex-vivo* tissue specimens/biopsies, only yields moderate inter-rater agreement. We hypothesize that deep learning can provide an automated tool to improve SCC grading, with potential to provide clinically relevant information *in-vivo* and *in-situ* using non-contact optical imaging. Previously in the literature, similar studies have been performed to predict

histological grade of other cancer types with other imaging modalities, such as MRI and digitized histology, with acceptable values of correlation agreements [111, 297, 298].

The objective of this project is to determine the ability of label-free, non-contact HSI for automatically predicting the histological grade of SCC in gross-level, *ex-vivo* tissue specimens using deep convolutional neural networks. The preliminary results of this work suggest that the combination of HSI and deep learning has the potential to produce near real-time SCC grading labelling, and future investigation is warranted.

A.2 Materials and Methods

Patients with head and neck squamous cell carcinoma (SCC) were recruited for this study after written and informed consent, conducted at Emory University School of Medicine, in collaboration with the Emory University Hospital Midtown (EUHM) departments of surgical pathology and otolaryngology [33, 113, 179]. Patient information was de-identified by a research coordinator at EUHM. For this investigation, we included only specimens of conventional, keratinizing SCC without HPV, all histological grades from well to poorly differentiated, and acquired tissue specimens of both the primary cancer and the tumor-involved cancer margin. The size of tissue specimens was approximately 10 mm in width and from 10 to 20 mm in height, by 2 to 3 mm in depth. Tissue specimens were acquired for research purposes from the surgical pathology department after being grossly examined, and the remaining tissue was used for the clinical pathology service.

The *ex-vivo* tissue specimens were imaged with HSI at an off-site imaging laboratory, after being transported in cold, buffered saline solution. After optical imaging,

specimens were fixed in formalin, paraffin embedded, microtome sectioned, and underwent H&E staining to produce histological glass slides. The H&E slides were used to obtain the ground truth for the gross, *ex-vivo* tissue specimens [33, 113, 179]. For specimens of the tumor-involved cancer margin, the cancerous SCC area was segmented and registered according to a previously reported protocol [184]. The histological ground truth was annotated by a board-certified pathologist with expertise in head and neck SCC.

This SCC grading study included 73 conventional SCC tissue specimens from all 56 patients undergoing surgical SCC resection with histological grade. The origin sites for the primary cancer were from the upper aerodigestive tract. A summary of the SCC grades and groups is shown in Table A-1. Because of insufficient tissue size to perform image patch making, described below, 10 tissue specimens were excluded from this study: not reported in Table A-1. The histological grade of SCC differentiation was obtained from the de-identified pathology report used for EUHM clinical service. There was a relative excess of grade 2 compared to 1 and 3, so only one primary cancer specimen was used from each patient with grade 2, moderately differentiated SCC.

Table A-1. Head and neck SCC data and samples, organized by histological grade of differentiation. Groups are shown as number of tissue specimens (number of patients).

SCC Histological Grade	Train	Validation	Test	Total
Well-differentiated (G1)	7 (4)	2 (1)	2 (1)	11 (6)
Moderately differentiated (G2)	30 (28)	5 (4)	4 (4)	39 (36)
Poorly differentiated (G3)	14 (9)	5 (3)	4 (2)	23 (14)
Total	51 (41)	12 (8)	10 (7)	73 (56)

Hyperspectral images were acquired using a Maestro imaging system (Perkin Elmer Inc., Waltham, Massachusetts). The imaging system consists of a Xenon white-light illumination source, a liquid crystal tunable filter to separate spectral bands, and a 16-bit charge-coupled device that acquires high-resolution images, 1,040 by 1,392 pixels, with a specimen-level resolution of approximately 25 microns per pixel [33, 113, 179]. The HSI were captured throughout the broad-band spectral range from 450 to 900 nm in 5 nm spectral increments, producing a hypercube with 91 spectral bands. The HSI data was calibrated and normalized using a standardized white reference disk and subtracting the inherent dark current of the sensor to obtain arbitrary units of reflectance [33, 113, 179]. There was substantial specular glare on some tissue specimens; this was left intact in the HSI, but reflectance values were capped at an intensity of 1.

To perform automatic SCC grading into three categories using HSI, a convolutional neural network (CNN) was used to train and classify the 3 SCC grades, validated on 12 tissues from 8 patients, and tested on 10 tissues from 7 patients to determine final performance. The CNN was implemented in TensorFlow [192] in Python on an Ubuntu machine with an i5 processor and NVIDIA Titan-XP GPU. The CNN architecture was originally inspired by the Inception-V4 CNN [217], but substantial modifications were made and squeeze-and-excitation modules were added. The input for the CNN was 3D HSI patches of size 101×101×91 (height, width, spectral bands). Patches were augmented by flips and rotations to be balanced across the three classes.

The CNN architecture was customized due to the uniqueness of the HSI input data and differing patch size from the original Inception-V4 architecture [217]. The customized CNN consisted of two convolutional layers, one max-pooling layer, and a final strided

convolutional layer with dropout to reduce the size to 25×25 before the inception modules. Only one inception-v4 A module was used, followed by 3 squeeze-and-excitation layers, before the reduction A module. One inception-v4 B module was used, followed by 6 squeeze-and-excitation layers, before the reduction B module. Lastly, one inception-v4 C module was used, followed by 2 squeeze-and-excitation layers, which was output to the logits by global average pooling. The adaptive delta gradient descent optimizer, AdaDelta, used to for training and weight optimization. The CNN was trained for 35 epochs of training data, using a batch size of 24 and learning rate of 0.01. The validation performance of all validation patches was calculated every 5 epochs, and training was stopped when patch-level performance peaked. The training time for one epoch was about 2 hours, and performing the testing on a series of patches was on the order of tens of seconds. After stopping training, tissue-level validation performance was calculated by taking the maximum probability of belonging to a SCC grade across all patches produced from a tissue and assigning that grade label to the entire tissue. Assigning tissue-level SCC grade using the maximum probability of all patches was also used for testing results. The rationale for this technique is that SCC grade is determined by the most aggressive grade present in a cancer specimen; therefore, it does not have to be uniform across and entire cancer tissue.

The CNN classification performance was evaluated using an external validation group and a fully-independent testing group of tissue specimens and patients. To evaluate the performance, the metrics of tissue level percent accuracy, Cohen's kappa statistic, and predictive probability were implemented. Cohen's kappa statistic (κ) for is a commonly used metric for determining agreement or concordance in ranking tasks, and the kappa

values ranges from [-1, +1], where 0 corresponds to randomly guessing. The predictive probability (P_k) was selected because it converges to area under the curve (AUC) for binary, continuous problems, so it offers relatability [299]. Values for predictive probability range from 0 to 1, where 0.5 corresponds to randomly guessing, just like AUC values.

The labels for the training data and the labels for the validation and testing groups was the SCC grade obtained from the EUHM clinical service. It is possible that there is some uncertainty in the grade labels because the labels were obtained from the entire clinical service gross tissue, not only the small specimens used for this research study. Additionally, there may be some inter-reader variability in the SCC grades labelled.

A.3 Results and Discussion

The overall multiclass performance on a tissue level for the validation group was 67% accuracy, 0.81 predictive probability, and a Cohen's kappa of 0.50 agreement. Within the validation group, one out of two grade 1 SCC were predicted correctly, with one incorrectly predicted as grade 2. Four out of five, grade 2 were predicted correctly, with one incorrectly predicted as grade 1. Three out of five, grade 3 were predicted correctly, with two incorrectly predicted as grade 1.

The tissue level multiclass performance for the testing group (10 tissues from 7 patients) was 60% accuracy, 0.70 predictive probability, and a Cohen's kappa of 0.44 agreement. One out of two grade 1 SCC were predicted correctly, with one incorrectly predicted as grade 2. All four grade 2 were predicted correctly. One out of four, grade 3

were predicted correctly, with three incorrectly predicted as grade 1. Table A-2 shows the summarized results for validation and testing groups.

Table A-2. Performance of HSI using CNN for automatic SCC grading, reported with predictive probability (P_k), percent accuracy, and Cohen's kappa (κ) calculated on a tissue specimen level.

HSI vs. Clinical Service	**Tissues (Patients)**	**P_k**	**Acc.**	**κ**
Validation	12 (8)	0.81	67%	0.50
Testing	10 (7)	0.70	60%	0.44

According to interpretation values for kappa, the validation and testing results represent fair to moderate agreement (0.44-0.50), while the inter-rater agreement represents borderline moderate to substantial agreement (0.60-0.64) [300]. This indicates that the results are promising, but the CNN training could benefit from more patient data, particularly grades 1 and 3, where we did not have enough samples. In the literature, automated histological grade has been performed directly using digitized histology with high kappas for different cancers: prostate cancer (0.71-0.75) [301], lung cancer (0.53) [284], and breast cancer (0.64) [302]. It is evident that maximum agreement, even using histology directly, differs by cancer type. Our results comparatively range from 0.44 to 0.50 and use gross-level tissue information only.

There a several limitations of this study. Firstly, due to a limited sample size, the statistical power of the results cannot be confidently stated. Moreover, with the limited number of specimens for grade 1 and 3, it can be seen that our proposed CNN method may be prone to over-predicting grade 2 carcinoma. Because of the over-abundance of grade 2 tissue specimens, it was possible to only use specimens of the primary tumor for grade 2.

However, grade 1 and 3 required including both primary tumor tissues and cancer-involved margin tissues that need to be registered and have the normal components segmented out.

It is possible that there could be disagreement between independent pathologist graders for SCC, as the literature suggests. With only one set of SCC grades for this study, we cannot identify the bounds of grade ranking disagreement for this sample size. A potential reason for the disagreement in SCC grading of the HSI predicted labels and the results of the EUHM clinical service could be that the tissue specimens acquired were not the exact grade of the entire gross-level tissue removed. This would confound the results and yield artificially lower evaluation metrics. The solution to this problem would be having several independent pathologist graders assign scores to the testing and validation groups for comparison. Nonetheless, the promising preliminary results demonstrate fair agreement of HSI predicted grade and pathologist assigned SCC grade.

We consider this study to be innovative because it designs and validates a quantitative, non-contact, and label-free approach for determining histological grade using optical imaging and deep learning. This method could have potential in helping surgeons and pathologist identify the aggressiveness of SCC either in the operating room or in the surgical pathology department without the need for histological sample preparation. The promising preliminary results warrant future investigation into the comparison of the automated methods as performed with digitized histology of the same specimens.

A.4 Conclusion

In conclusion, this addendum section reports the preliminary results from HSI for automatically predicting SCC grade of head and neck gross-level tissue specimens. This

work is the first of its kind to explore HSI for the automatic grading of SCC gross tissues, and predictions from the proposed HSI and CNN method agree with the SCC grade from clinical service to a fair to moderate degree. The clinical problem of SCC grading is a challenging task with only moderate inter-rater agreement. This warrants future investigation into the comparison of the automated methods performed with digitized histology of the same specimens to determine the agreement of grade predictive algorithms with different imaging modalities. This addendum to the thesis work further demonstrates the potential of HSI for clinical application.

APPENDIX B. SELECTED PUBLICATIONS

First-Author Journal Publications:

1. Ortega, S., **Halicek, M.**, et al. and Fei, B., "Hyperspectral imaging for the detection of glioblastoma tumor cells in H&E slides using convolutional neural networks," Sensors (Submitted). ***Co-first authorship**
2. Ortega, S., **Halicek, M.**, et al. and Fei, B., "The use of hyperspectral and multispectral imaging in digital histology: A systematic review," Biomedical Optics Express (Submitted). ***Co-first authorship**
3. **Halicek, M.**, Dormer, J., et al. and Fei, B., "Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning," Biomedical Optics Express 11(9), 1367 (2020).
4. **Halicek, M.**, Shahedi, M., et al. and Fei, B., "Detection of head and neck cancers in digitized whole-slide histological images using convolutional neural networks," Scientific Reports 9(1) 1-11 (2019).
5. **Halicek, M.**, Dormer, J., et al. and Fei, B., "Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning," Cancers 11(9), 1367 (2019).
6. **Halicek, M.**, Fabelo, H., et al. and Fei, B., "Hyperspectral imaging for head and neck cancer detection: specular glare and variance of the tumor margin in surgical specimens," J Medical Imaging 6(3), 035004 (2019).
7. **Halicek, M.**, Fabelo, H., et al. and Fei, B., "In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: revealing the invisible features of cancer," Cancers 11(6), 756 (2019).
8. Fabelo, H., **Halicek, M.**, et al., and Fei, B. "Deep learning-based framework for in-vivo identification of glioblastoma tumor using hyperspectral images of human brain," Sensors 19(4), 920 (2019). ***Co-first authorship**
9. **Halicek, M.**, Little, J. V., Chen, A. Y., and Fei, B., "Optical biopsy of head and neck cancers using hyperspectral imaging and convolutional neural networks," J Biomed Optics 24(3), 036007 (2019).
10. **Halicek, M.**, Lu, G., Little, J. V., Wang, X., et al, and Fei, B., "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," J Biomed Optics 22(6), 060503 (2017).

First-Author Conference Proceedings:

1. **Halicek, M.**, et al., and Fei, B. "Conditional generative adversarial network for synthesizing hyperspectral images of breast cancer cells from digitized RGB histology," Proc. SPIE, Medical Imaging: Digital Pathology. 109560K (2020).
2. **Halicek, M.**, et al., and Fei, B., "Cancer Margin Evaluation using Machine Learning and Hyperspectral Imaging of Head and Neck Cancer," Proc. SPIE, Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling. 109511A (2019).
3. **Halicek, M.**, et al., and Fei, B., "Detection of squamous cell carcinoma in digitized histological images from the head and neck using convolutional neural networks," Proc. SPIE, Medical Imaging: Digital Pathology. 109560K (2019).
4. Fabelo, H. **Halicek, M.**, et al., and Fei, B., "Surgical aid visualization system for glioblastoma tumor identification based on deep learning techniques using in-vivo human brain hyperspectral images," Proc. SPIE, Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling. 1095110 (2019). ***Co-first authorship**
5. **Halicek, M.**, Little, J. V., Wang, X., Patel, M., Griffith, C. C., El-Deiry, M. W., Chen, A. Y., and Fei, B., "Tumor Margin Classification of Head and Neck Cancer Using Hyperspectral Imaging and Convolutional Neural Networks," Proc. SPIE, Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling. 1057605 (2018).
6. **Halicek, M.**, Little, J. V., Wang, X., Patel, M., Griffith, C. C., El-Deiry, M. W., Chen, A. Y., and Fei, B., "Deformable registration of histological cancer margins to gross hyperspectral images using demons," Proc. SPIE, Medical Imaging: Digital Pathology. 105810N (2018).
7. **Halicek, M.**, Little, J. V., Wang, X., Patel, M., Griffith, C. C., El-Deiry, M. W., Chen, A. Y., and Fei, B., "Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks," Proc. SPIE, Photonics West: Optical Imaging, Therapeutics, and Adv. Technology in Head & Neck Surgery and Otolaryngology 104690X (2018).

Conference Oral Presentations:

1. **Halicek, M.**, et. al. "Conditional generative adversarial network for synthesizing hyperspectral images of breast cancer cells from digitized RGB histology," SPIE Medical Imaging Conf. 2020. Houston, TX (Feb. 2020).

2. **Halicek, M.**, et. al. "Hyperspectral Imaging for Classification of Histologic Grade in Gross Specimens of Squamous Cell Carcinoma Using Deep Learning," AAPM 61st Annual Meeting, San Antonio, TX (July 2019).

3. **Halicek, M.**, et. al. "Detection of squamous cell carcinoma in digitized histological images from the head and neck using convolutional neural networks," SPIE Medical Imaging Conf. 2019. San Diego, CA (Feb. 2019).

4. **Halicek, M.**, et. al. "Cancer Margin Evaluation using Machine Learning and Hyperspectral Imaging of Head and Neck Cancer," SPIE Medical Imaging Conference 2019. San Diego, CA (Feb. 2019).

5. **Halicek, M.**, et. al. "Automated Detection of Head and Neck Squamous Cell Carcinoma Digitized Whole-slide Histological Images using Deep Learning," Ann. Meeting of Biomedical Eng. Society Atlanta, GA (Oct. 2018).

6. **Halicek, M.**, et. al. "Tumor Margin Classification of Head and Neck Cancer Using Hyperspectral Imaging and Convolutional Neural Networks," SPIE Medical Imaging Conference 2018. Houston, TX (Feb. 2018).

7. **Halicek, M.**, et. al. "Deformable registration of histological cancer margins to gross hyperspectral images using demons," SPIE Medical Imaging Conference 2018. Houston, TX (Feb. 2018).

8. **Halicek, M.**, et. al. "Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks," SPIE Photonics West Conference 2018. San Francisco, CA (Jan. 2018).

Other Publications:

1. Shahedi, M., **Halicek, M.**, et al. and Fei, B., "Deep learning-based 3D segmentation of the prostate on CT images," J Medical Imaging 6(2), 25003 (2019).
2. Shahedi, M., **Halicek, M.**, Guo, R., Zhang, G., Schuster, D., and Fei, B., "A semiautomatic segmentation method for prostate in CT images using local texture classification and statistical shape modeling," Med Phys 45(6), 2527-41 (2018).
3. Ling, M., **Halicek, M**, et al., and Fei, B. "Hyperspectral microscopic imaging for histological head and neck squamous cell carcinoma detection using machine learning," Proc. SPIE, Medical Imaging: Digital Pathology. 11320 (2020).
4. Ling, M., **Halicek, M**, et al., and Fei, B. "In vivo cancer detection in animal model using hyperspectral image classification with wavelet feature extraction," Proc. SPIE, Medical Imaging: Biomedical Applications in Molecular, Structural, and Functional Imaging. 11317 (2020).

5. Ortega, S., **Halicek, M**, et al., and Fei, B. "Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized pathologic images," Proc. SPIE, Medical Imaging: Digital Pathology. 11320 (2020).
6. Shahedi, M., **Halicek, M**, et al., and Fei, B. "Incorporating minimal user input into deep learning based image segmentation," Proc. SPIE, Medical Imaging: Computer-Aided Diagnosis. 11314 (2020).
7. Mavuduru, A., **Halicek, M**, et al., and Fei, B. "Using a 22-layer U-Net to perform segmentation of squamous cell carcinoma on digitized head and neck histological images," Proc. SPIE, Medical Imaging: Digital Pathology. 11320 (2020).
8. Tran, C., **Halicek, M**, et al., and Fei, B. "Fully automated segmentation of the right ventricle in patients with repaired Tetralogy of Fallot using U-Net," Proc. SPIE, Medical Imaging: Biomedical Applications in Molecular, Structural, and Functional Imaging. 11317 (2020).
9. Huang, J., **Halicek, M**, et al., and Fei, B. "Augmented reality visualization of hyperspectral imaging classifications for image-guided brain tumor resection," Proc. SPIE, Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling. 11315 (2020).
10. Zhou, X., **Halicek, M**, et al., and Fei, B. "Development of a new polarized hyperspectral imaging microscope," Proc. SPIE, Photonics West: Imaging, Therapeutics, & Adv. Technology in H&N Surgery,Otolaryngology. 11213 (2020).
11. Shahedi, M., **Halicek, M.**, et al., and Fei, B., "A semiautomatic approach for prostate segmentation in MR images using local texture classification and statistical shape modeling," Proc. SPIE, Image-Guided Procedures, Robotic Interventions, and Modeling 109512I (2019).
12. Shahedi, M., Ma, L., **Halicek, M.,** Guo, R., Zhang, G., Schuster, D.M., Nieh, P.T., Master, V.V., and Fei, B., "A semiautomatic algorithm for three-dimensional segmentation of the prostate on CT images using shape and local texture characteristics," Proc. SPIE, Image-Guided Procedures, Robotic Interventions, and Modeling 1057616 (2018).
13. Dormer, J., Ma, L., **Halicek, M.**, Reilly, C.M., Schreibmann, E., and Fei, B., "Heart chamber segmentation from CT using convolutional neural networks," Proc. SPIE, Biomed. App. in Molecular, Structural, and Functional Imaging 105782S (2018).
14. Dormer, J., **Halicek, M.**, Ma, L., Reilly, C.M., Schreibmann, E., and Fei, B., "Convolutional neural networks for the detection of diseased hearts using CT images," Proc. SPIE 10575, Computer-Aided Diagnosis 1057530 (2018).
15. B. Fei, Lu, G., **Halicek, M.**, Little, J. V., Wang, X., Patel, M., Griffith, C. C., El-Deiry, M. W., Chen, A. Y., "Label-free hyperspectral imaging and quantification methods for surgical margin assessment of tissue specimens of cancer patients," 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, p. 4041-45 (2017).

REFERENCES

1. S. I. Ferlay J, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F., "GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11," (International Agency for Research on Cancer, Lyon, France, 2013).

2. L. J. Joseph, M. Goodman, K. Higgins, R. Pilai, S. S. Ramalingam, K. Magliocca, M. R. Patel, M. El-Deiry, J. T. Wadsworth, T. K. Owonikoko, J. J. Beitler, F. R. Khuri, D. M. Shin, and N. F. Saba, "Racial disparities in squamous cell carcinoma of the oral tongue among women: a SEER data analysis," Oral Oncol **51**, 586-592 (2015).

3. A. Argiris, M. V. Karamouzis, D. Raben, and R. L. Ferris, "Head and neck cancer," The Lancet **371**, 1695-1709 (2008).

4. B. Y. Kim, J. E. Choi, E. Lee, Y. I. Son, C. H. Baek, S. W. Kim, and M. K. Chung, "Prognostic factors for recurrence of locally advanced differentiated thyroid cancer," J Surg Oncol **116**(2017).

5. J. D. Meier, D. A. Oliver, and M. A. Varvares, "Surgical margin determination in head and neck oncology: current clinical practice. The results of an International American Head and Neck Society Member Survey," Head Neck **27**, 952-958 (2005).

6. M. L. Hinni, A. Ferlito, M. S. Brandwein-Gensler, R. P. Takes, C. E. Silver, W. H. Westra, R. R. Seethala, J. P. Rodrigo, J. Corry, C. R. Bradford, J. L. Hunt, P. Strojan, K. O. Devaney, D. R. Gnepp, D. M. Hartl, L. P. Kowalski, A. Rinaldo, and L. Barnes, "Surgical margins in head and neck cancer: A contemporary review," Head & Neck **35**, 1362-1370 (2013).

7. D. A. Novis and R. J. Zarbo, "Interinstitutional comparison of frozen section turnaround time. A College of American Pathologists Q-Probes study of 32868 frozen sections in 700 hospitals," Archives of pathology & laboratory medicine **121**, 559-567 (1997).

8. T. Y. Chen, L. J. Emrich, and D. L. Driscoll, "The clinical significance of pathological findings in surgically resected margins of the primary tumor in head and neck carcinoma," International Journal of Radiation Oncology • Biology • Physics **13**, 833-837 (1987).

9. T. R. Loree and E. W. Strong, "Significance of positive margins in oral cavity squamous carcinoma," The American Journal of Surgery **160**, 410-414 (1990).

10. D. N. Sutton, J. S. Brown, S. N. Rogers, E. D. Vaughan, and J. A. Woolgar, "The prognostic implications of the surgical margin in oral squamous cell carcinoma," International Journal of Oral and Maxillofacial Surgery **32**, 30-34 (2003).

11. S. Rathod, J. Livergant, J. Klein, I. Witterick, and J. Ringash, "A systematic review of quality of life in head and neck cancer treated with surgery with or without adjuvant treatment," Oral Oncol **51**, 888-900 (2015).

12. T. M. Pimenta Amaral, A. R. da Silva Freire, A. L. Carvalho, C. A. L. Pinto, and L. P. Kowalski, "Predictive factors of occult metastasis and prognosis of clinical stages I and II squamous cell carcinoma of the tongue and floor of the mouth," Oral Oncology **40**, 780-786 (2004).

13. P. Garzino-Demo, A. Dell'Acqua, P. Dalmasso, M. Fasolis, G. M. La Terra Maggiore, G. Ramieri, S. Berrone, M. Rampino, and M. Schena, "Clinicopathological parameters and outcome of 245 patients operated for oral squamous cell carcinoma," Journal of Cranio-Maxillo-Facial Surgery **34**, 344-350 (2006).

14. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," Journal of biomedical optics **19**, 10901 (2014).

15. D. Ravi, H. Fabelo, G. M. Callico, and G. Yang, "Manifold Embedding and Semantic Segmentation for Intraoperative Guidance with Hyperspectral Brain Imaging," IEEE Trans Med Imaging **36**, 1845 (2017).

16. A. Pardo, E. Real, V. Krishnaswamy, J. M. Lopez-Higuera, B. W. Pogue, and O. M. Conde, "Directional Kernel Density Estimation for Classification of Breast Tissue Spectra," IEEE Trans Med Imaging **36**, 64-73 (2017).

17. M. Milanic, L. A. Paluchowski, and L. L. Randeberg, "Hyperspectral imaging for detection of arthritis: feasibility and prospects," Journal of biomedical optics **20**, 096011 (2015).

18. S. V. Panasyuk, S. Yang, D. V. Faller, D. Ngo, R. A. Lew, J. E. Freeman, and A. E. Rogers, "Medical hyperspectral imaging to facilitate residual tumor identification during surgery," Cancer Biology & Therapy **6**, 439-446 (2014).

19. M. E. Martin, M. B. Wabuyele, K. Chen, P. Kasili, M. Panjehpour, M. Phan, B. Overholt, G. Cunningham, D. Wilson, R. C. Denovo, and T. Vo-Dinh, "Development of an advanced hyperspectral imaging (HSI) system with applications for cancer detection," Ann Biomed Eng **34**, 1061-1068 (2006).

20. P. Liu, J. Huang, S. Zhang, and R. X. Xu, "Multiview hyperspectral topography of tissue structural and functional characteristics," Journal of biomedical optics **21**, 16012 (2016).

21. H. T. Lim and V. M. Murukeshan, "A four-dimensional snapshot hyperspectral video-endoscope for bio-imaging applications," Sci Rep **6**, 24044 (2016).

22. S. J. Leavesley, M. Walters, C. Lopez, T. Baker, P. F. Favreau, T. C. Rich, P. F. Rider, and C. W. Boudreaux, "Hyperspectral imaging fluorescence excitation scanning for colon cancer detection," Journal of biomedical optics **21**, 104003 (2016).

23. R. T. Kester, N. Bedard, L. Gao, and T. S. Tkaczyk, "Real-time snapshot hyperspectral imaging endoscope," Journal of biomedical optics **16**, 056005 (2011).

24. L. Gao and R. T. Smith, "Optical hyperspectral imaging in microscopy and spectroscopy - a review of data acquisition," J Biophotonics **8**, 441-456 (2015).

25. F. Blanco, F. Lumbreras, J. Serrat, R. Siener, S. Serranti, G. Bonifazi, M. Lopez-Mesas, and M. Valiente, "Taking advantage of hyperspectral imaging classification of urinary stones against conventional infrared spectroscopy," Journal of biomedical optics **19**, 126004 (2014).

26. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, 1097-1105 (2012).

27. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," arXiv: Computational Research Repository **1409.4842**(2014).

28. V. Slavkovikj, S. Verstockt, W. De Neve, S. Van Hoecke, and R. Van de Walle, "Hyperspectral Image Classification with Convolutional Neural Networks," Proceedings of the 23rd ACM international conference on Multimedia, 1159-1162 (2015).

29. D. Yudovsky, A. Nouvong, K. Schomacker, and L. Pilon, "Assessing diabetic foot ulcer development risk with hyperspectral tissue oximetry," Journal of biomedical optics **16**, 026009 (2011).

30. G. Lu, L. Halig, D. Wang, X. Qin, Z. G. Chen, and B. Fei, "Spectral-spatial classification for noninvasive cancer detection using hyperspectral imaging," Journal of biomedical optics **19**, 106004 (2014).

31. G. Lu, X. Qin, D. Wang, Z. G. Chen, and B. Fei, "Quantitative Wavelength Analysis and Image Classification for Intraoperative Cancer Diagnosis with Hyperspectral Imaging," Proc SPIE Int Soc Opt Eng **9415**(2015).

32. G. Lu, D. Wang, X. Qin, S. Muller, X. Wang, A. Y. Chen, Z. G. Chen, and B. Fei, "Detection and delineation of squamous neoplasia with hyperspectral imaging in a mouse model of tongue carcinogenesis," J Biophotonics (2017).

33. G. Lu, J. V. Little, X. Wang, H. Zhang, M. Patel, C. C. Griffith, M. El-Deiry, A. Y. Chen, and B. Fei, "Detection of Head and Neck Cancer in Surgical Specimens Using Quantitative Hyperspectral Imaging," Clin Cancer Res **23**, 5426-5438 (2017).

34. M. Kamruzzaman and D. W. Sun, "Introduction to Hyperspectral Imaging Technology," in *Computer Vision Technology for Food Quality Evaluation,* (Elsevier, 2016), pp. 111-139.

35. C. Starr, *Biology : concepts and applications / Cecie Starr* (Wadsworth Pub., 1997).

36. D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," IEEE Signal Processing Magazine **19**, 29-43 (2002).

37. M. Govender, K. Chetty, and H. Bulcock, "A review of hyperspectral remote sensing and its application in vegetation and water resource studies," Water SA **33**(2009).

38. F. D. van der Meer, H. M. A. van der Werff, F. J. A. van Ruitenbeek, C. A. Hecker, W. H. Bakker, M. F. Noomen, M. van der Meijde, E. J. M. Carranza, J. B. d. Smeth, and T. Woldai, "Multi- and hyperspectral geologic remote sensing: A review," International Journal of Applied Earth Observation and Geoinformation **14**, 112-128 (2012).

39. H. Liang, "Advances in multispectral and hyperspectral imaging for archaeology and art conservation," Applied Physics A **106**, 309-323 (2011).

40. V. Miljković and D. Gajski, "Adaptation of Industrial Hyperspectral Line Scanner for Archaeological Applications," ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **XLI-B5**, 343-345 (2016).

41. W. Fortunato de Carvalho Rocha, G. P. Sabin, P. H. Março, and R. J. Poppi, "Quantitative analysis of piroxicam polymorphs pharmaceutical mixtures by hyperspectral imaging and chemometrics," Chemometrics and Intelligent Laboratory Systems **106**, 198-204 (2011).

42. L. d. M. França, M. F. Pimentel, S. d. S. Simões, S. Grangeiro, J. M. Prats-Montalbán, and A. Ferrer, "NIR hyperspectral imaging to evaluate degradation in captopril commercial tablets," European Journal of Pharmaceutics and Biopharmaceutics **104**, 180-188 (2016).

43. G. J. Edelman, E. Gaston, T. G. van Leeuwen, P. J. Cullen, and M. C. G. Aalders, "Hyperspectral imaging for non-contact analysis of forensic traces," Forensic Science International **223**, 28-39 (2012).

44. C. S. Silva, M. F. Pimentel, R. S. Honorato, C. Pasquini, J. M. Prats-Montalbán, and A. Ferrer, "Near infrared hyperspectral imaging for forensic analysis of document forgery," The Analyst **139**, 5176-5184 (2014).

45. M. Á. Fernández de la Ossa, J. M. Amigo, and C. García-Ruiz, "Detection of residues from explosive manipulation by near infrared hyperspectral imaging: A promising forensic tool," Forensic Science International **242**, 228-235 (2014).

46. P. W. Yuen and M. Richardson, "An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition," The Imaging Science Journal **58**, 241-253 (2010).

47. V. C. Coffey, "Hyperspectral Imaging for Safety and Security," Optics and Photonics News **26**, 26 (2015).

48. M. Teke, H. S. Deveci, O. Haliloglu, S. Z. Gurbuz, and U. Sakarya, "A short survey of hyperspectral remote sensing applications in agriculture," in *2013 6th International Conference on Recent Advances in Space Technologies (RAST),* (IEEE, 2013).

49. L. M. Dale, A. Thewis, C. Boudry, I. Rotar, P. Dardenne, V. Baeten, and J. A. F. Pierna, "Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review," Applied Spectroscopy Reviews **48**, 142-159 (2013).

50. D. Wu and D.-W. Sun, "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals," Innovative Food Science & Emerging Technologies **19**, 1-14 (2013).

51. Y.-Z. Feng and D.-W. Sun, "Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review," Critical Reviews in Food Science and Nutrition **52**, 1039-1058 (2012).

52. D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O. L. García-Navarrete, and J. Blasco, "Recent Advances and Applications of Hyperspectral Imaging for Fruit and Vegetable Quality Assessment," Food and Bioprocess Technology **5**, 1121-1142 (2011).

53. M. A. Calin, S. V. Parasca, D. Savastru, and D. Manea, "Hyperspectral Imaging in the Medical Field: Present and Future," Applied Spectroscopy Reviews **49**, 435-447 (2014).

54. Q. Li, X. He, Y. Wang, H. Liu, D. Xu, and F. Guo, "Review of spectral imaging technology in biomedical engineering: achievements and challenges," Journal of biomedical optics **18**, 100901 (2013).

55. A. Bjorgan, M. Denstedt, M. Milanič, L. A. Paluchowski, and L. L. Randeberg, "Vessel contrast enhancement in hyperspectral images," in *Optical Biopsy XIII: Toward Real-Time Spectroscopic Imaging and Diagnosis,* (SPIE, 2015).

56. H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Blood vessel detection and artery-vein differentiation using hyperspectral imaging," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* (IEEE, 2009).

57. H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging," IEEE Trans Biomed Eng **57**, 2011-2017 (2010).

58. D. J. Mordant, I. Al-Abboud, G. Muyo, A. Gorman, A. Sallam, P. Ritchie, A. R. Harvey, and A. I. McNaught, "Spectral imaging of the retina," Eye **25**, 309-320 (2011).

59. W. R. Johnson, D. W. Wilson, W. Fink, M. Humayun, and G. Bearman, "Snapshot hyperspectral imaging in ophthalmology," Journal of biomedical optics **12**, 014036 (2007).

60. L. Gao, R. T. Smith, and T. S. Tkaczyk, "Snapshot hyperspectral retinal camera with the Image Mapping Spectrometer (IMS)," Biomedical Optics Express **3**, 48 (2011).

61. M. Milanic, A. Bjorgan, M. Larsson, T. Strömberg, and L. L. Randeberg, "Detection of hypercholesterolemia using hyperspectral imaging of human skin," in *Clinical and Biomedical Spectroscopy and Imaging IV,* (SPIE, 2015).

62. L. Zhi, D. Zhang, J.-q. Yan, Q.-L. Li, and Q.-l. Tang, "Classification of hyperspectral medical tongue images for tongue diagnosis," Computerized Medical Imaging and Graphics **31**, 672-678 (2007).

63. D. Yudovsky, A. Nouvong, and L. Pilon, "Hyperspectral Imaging in Diabetic Foot Wound Care," Journal of Diabetes Science and Technology **4**, 1099-1113 (2010).

64. T. Adão, J. Hruška, L. Pádua, J. Bessa, E. Peres, R. Morais, and J. Sousa, "Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry," Remote Sensing **9**, 1110 (2017).

65. N. Hagen and M. W. Kudenov, "Review of snapshot spectral imaging technologies," Optical Engineering **52**, 090901 (2013).

66. M. Li, S. Zang, B. Zhang, S. Li, and C. Wu, "A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information," European Journal of Remote Sensing **47**, 389-411 (2014).

67. M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in Spectral-Spatial Classification of Hyperspectral Images," Proceedings of the IEEE **101**, 652-675 (2013).

68. G. Cano, J. Garcia-Rodriguez, A. Garcia-Garcia, H. Perez-Sanchez, J. A. Benediktsson, A. Thapa, and A. Barr, "Automatic selection of molecular descriptors using random forest: Application to drug discovery," Expert Systems with Applications **72**, 151-159 (2017).

69. T. Puchert, D. Lochmann, J. C. Menezes, and G. Reich, "Near-infrared chemical imaging (NIR-CI) for counterfeit drug identification—A four-stage concept with a novel approach of data processing (Linear Image Signature)," Journal of Pharmaceutical and Biomedical Analysis **51**, 138-145 (2010).

70. A. Gowen, C. Odonnell, P. Cullen, G. Downey, and J. Frias, "Hyperspectral imaging – an emerging process analytical tool for food quality and safety control," Trends in Food Science & Technology **18**, 590-598 (2007).

71. Y. Montembeault, P. Lagueux, V. Farley, A. Villemaire, and K. C. Gross, "Hyper-Cam: Hyperspectral IR imaging applications in defence innovative research," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, (IEEE, 2010).

72. P. Fisher, "The pixel: A snare and a delusion," International Journal of Remote Sensing **18**, 679-685 (1997).

73. C.-I. Chang, "Hyperspectral Measures for Spectral Characterization," in *Hyperspectral Imaging*, (Springer US, 2003), pp. 15-35.

74. H. Grahn, "Techniques and Applications of Hyperspectral Image Analysis," (John Wiley & Sons, Ltd, 2007).

75. T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Multiple Classifier Systems*, (Springer Berlin Heidelberg, 2000), pp. 1-15.

76. J. C.-W. Chan and D. Paelinckx, "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," Remote Sensing of Environment **112**, 2999-3011 (2008).

77. J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," IEEE Transactions on Geoscience and Remote Sensing **43**, 480-491 (2005).

78. H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Wavelet-Based Compression and Segmentation of Hyperspectral Images in Surgery," in *Lecture Notes in Computer Science*, (Springer Berlin Heidelberg), pp. 142-149.

79. F. Blanco, M. Lopez-Mesas, S. Serranti, G. Bonifazi, J. Havel, and M. Valiente, "Hyperspectral imaging based method for fast characterization of kidney stone types," Journal of biomedical optics **17**, 076027 (2012).

80. G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing **43**, 1351-1362 (2005).

81. Y.-D. Zhang, Z.-J. Yang, H.-M. Lu, X.-X. Zhou, P. Phillips, Q.-M. Liu, and S.-H. Wang, "Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine, and Stratified Cross Validation," IEEE Access **4**, 8375-8385 (2016).

82. Y. Zhang, S. Lu, X. Zhou, M. Yang, L. Wu, B. Liu, P. Phillips, and S. Wang, "Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine," SIMULATION **92**, 861-871 (2016).

83. S. Wang, M. Yang, S. Du, J. Yang, B. Liu, J. M. Gorriz, J. Ramírez, T.-F. Yuan, and Y. Zhang, "Wavelet Entropy and Directed Acyclic Graph Support Vector Machine for Detection of Patients with Unilateral Hearing Loss in MRI Scanning," Frontiers in Computational Neuroscience **10**(2016).

84. S. Theodoridis and K. Koutroumbas, "Clustering Algorithms I: Sequential Algorithms," in *Pattern Recognition*, (Elsevier, 2009), pp. 627-652.

85. A. Moore, "Hierarchical and k-Means Clustering," in *Discovering Knowledge in Data*, (John Wiley & Sons, Inc., 2014), pp. 209-227.

86. F. Zhang, B. Du, L. Zhang, and L. Zhang, "Hierarchical feature learning with dropout k-means for hyperspectral image classification," Neurocomputing **187**, 75-82 (2016).

87. S. Narumalani, D. R. Mishra, J. Burkholder, P. B. T. Merani, and G. Willson, "A Comparative Evaluation of ISODATA and Spectral Angle Mapping for the Detection of Saltcedar Using Airborne Hyperspectral Imagery," Geocarto International **21**, 59-66 (2006).

88. S. A. E. Rahman, "Hyperspectral Imaging Classification Using ISODATA Algorithm: Big Data Challenge," in *2015 Fifth International Conference on e-Learning (econf)*, (IEEE, 2015).

89. S. C. Johnson, "Hierarchical clustering schemes," Psychometrika **32**, 241-254 (1967).

90. N. Gillis, D. Kuang, and H. Park, "Hierarchical Clustering of Hyperspectral Images Using Rank-Two Nonnegative Matrix Factorization," IEEE Transactions on Geoscience and Remote Sensing **53**, 2066-2078 (2015).

91. A. R. Barakbah and K. Arai, "Centronit: Initial Centroid Designation Algorithm for K-Means Clustering," EMITTER International Journal of Engineering Technology **2**(2014).

92. A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, "Generative model-based clustering of directional data," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03,* (ACM Press, 2003).

93. K. Masood, N. Rajpoot, K. Rajpoot, and H. Qureshi, "Hyperspectral Colon Tissue Classification using Morphological Analysis," in *2006 International Conference on Emerging Technologies,* (IEEE, 2006).

94. B. Regeling, B. Thies, A. O. Gerstner, S. Westermann, N. A. Muller, J. Bendix, and W. Laffers, "Hyperspectral Imaging Using Flexible Endoscopy for Laryngeal Cancer Detection," Sensors (Basel) **16**(2016).

95. L. Zhang, L. Zhang, and B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," IEEE Geoscience and Remote Sensing Magazine **4**, 22-40 (2016).

96. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436-444 (2015).

97. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE **86**, 2278-2324 (1998).

98. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," Neural Computation **1**, 541-551 (1989).

99. W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep Convolutional Neural Networks for Hyperspectral Image Classification," Journal of Sensors **2015**, 1-12 (2015).

100. J. Yue, W. Zhao, S. Mao, and H. Liu, "Spectral–spatial classification of hyperspectral images using deep convolutional neural networks," Remote Sensing Letters **6**, 468-477 (2015).

101. W. Zhao and S. Du, "Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach," IEEE Transactions on Geoscience and Remote Sensing **54**, 4544-4554 (2016).

102. D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep Learning for Health Informatics," IEEE Journal of Biomedical and Health Informatics **21**, 4-21 (2017).

103. M. M. A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," Information Sciences **345**, 340-354 (2016).

104. Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," Journal of Neural Engineering **14**, 016003 (2016).

105. S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation," The Scientific World Journal **2014**, 1-10 (2014).

106. D. Maji, A. Santara, S. Ghosh, D. Sheet, and P. Mitra, "Deep neural network and random forest hybrid architecture for learning to detect retinal vessels in fundus images," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* (IEEE, 2015).

107. M. Melinscak, P. Prentasic, and S. Loncaric, "Retinal Vessel Segmentation using Deep Neural Networks," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications,* (SCITEPRESS - Science and and Technology Publications, 2015).

108. A. Lahiri, A. G. Roy, D. Sheet, and P. K. Biswas, "Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* (IEEE, 2016).

109. P.-J. Chen, M.-C. Lin, M.-J. Lai, J.-C. Lin, H. H.-S. Lu, and V. S. Tseng, "Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis," Gastroenterology **154**, 568-575 (2018).

110. E. Ribeiro, A. Uhl, G. Wimmer, and M. Häfner, "Exploring Deep Learning and Transfer Learning for Colonic Polyp Classification," Computational and Mathematical Methods in Medicine **2016**, 1-16 (2016).

111. D. Bychkov, N. Linder, R. Turkki, S. Nordling, P. E. Kovanen, C. Verrill, M. Walliander, M. Lundin, C. Haglund, and J. Lundin, "Deep learning based tissue analysis predicts outcome in colorectal cancer," Sci Rep **8**, 3395 (2018).

112. K. Sirinukunwattana, S. E. Ahmed Raza, T. Yee-Wah, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images," IEEE Trans Med Imaging **35**, 1196-1206 (2016).

113. M. Halicek, G. Lu, J. V. Little, X. Wang, M. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," Journal of biomedical optics **22**, 60503 (2017).

114. P.-C. Chen and W.-C. Lin, "Spectral-profile-based algorithm for hemoglobin oxygen saturation determination from diffuse reflectance spectra," Biomedical Optics Express **2**, 1082 (2011).

115. W. A. Eaton, L. K. Hanson, P. J. Stephens, J. C. Sutherland, and J. B. R. Dunn, "Optical spectra of oxy- and deoxyhemoglobin," Journal of the American Chemical Society **100**, 4991-5003 (1978).

116. S. K. Sekar, I. Bargigia, A. D. Mora, P. Taroni, A. Ruggeri, A. Tosi, A. Pifferi, and A. Farina, "Diffuse optical characterization of collagen absorption from 500 to 1700 nm," Journal of biomedical optics **22**, 15006 (2017).

117. S. L. Jacques, "Optical properties of biological tissues: a review," Physics in Medicine and Biology **58**, R37-R61 (2013).

118. J. Deal, S. J. McFarland, A. Robinson, A. Alford, D. Weber, T. Rich, and S. Leavesley, "Hyperspectral imaging fluorescence excitation scanning spectral characteristics of remodeled mouse arteries," in *Label-free Biomedical Imaging and Sensing (LBIS) 2019,* (SPIE, 2019).

119. B. H. Herrmann and C. Hornberger, "Monte-Carlo Simulation of Light Tissue Interaction in Medical Hyperspectral Imaging Applications," Current Directions in Biomedical Engineering **4**, 275-278 (2018).

120. R. H. Wilson and M. A. Mycek, "Models of Light Propagation in Human Tissue Applied to Cancer Diagnostics," Technology in Cancer Research & Treatment **10**, 121-134 (2011).

121. R. Lambert, "Endoscopy in screening for digestive cancer," World Journal of Gastrointestinal Endoscopy **4**, 518 (2012).

122. M. M. Laterza, M. Orditura, A. Fabozzi, J. Ventriglia, B. Savastano, A. Petrillo, G. Giordano, E. Martinelli, V. Gambardella, F. Ciardiello, and F. De Vita, "Increased VEGF-C serum levels are predictive of a poor outcome in patients with resectable gastric cancer," Annals of Oncology **25**(2014).

123. A. J. M. Watson and P. D. Collins, "Colon Cancer: A Civilization Disorder," Digestive Diseases **29**, 222-228 (2011).

124. M. Pascual, "Laparoscopic colorectal surgery: Current status and implementation of the latest technological innovations," World Journal of Gastroenterology **22**, 704 (2016).

125. M. V. Ottermo, M. vstedal, T. Lang, y. Stavdahl, Y. Yavuz, T. A. Johansen, and R. M??rvik, "The Role of Tactile Feedback in Laparoscopic Surgery," Surgical Laparoscopy, Endoscopy & Percutaneous Techniques **16**, 390-400 (2006).

126. H. Akbari, K. Uto, Y. Kosugi, K. Kojima, and N. Tanaka, "Cancer detection using infrared hyperspectral imaging," Cancer Sci **102**, 852-857 (2011).

127. S. Kiyotoki, J. Nishikawa, T. Okamoto, K. Hamabe, M. Saito, A. Goto, Y. Fujita, Y. Hamamoto, Y. Takeuchi, S. Satori, and I. Sakaida, "New method for detection of gastric cancer by hyperspectral imaging: a pilot study," Journal of biomedical optics **18**, 26010 (2013).

128. A. Goto, J. Nishikawa, S. Kiyotoki, M. Nakamura, J. Nishimura, T. Okamoto, H. Ogihara, Y. Fujita, Y. Hamamoto, and I. Sakaida, "Use of hyperspectral imaging technology to develop a diagnostic support system for gastric cancer," Journal of biomedical optics **20**, 016017 (2015).

129. E. J. M. Baltussen, E. N. D. Kok, S. G. Brouwer de Koning, J. Sanders, A. G. J. Aalbers, N. F. M. Kok, G. L. Beets, C. C. Flohil, S. C. Bruin, K. F. D. Kuhlmann, H. Sterenborg, and T. J. M. Ruers, "Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery," Journal of biomedical optics **24**, 1-9 (2019).

130. Z. Han, A. Zhang, X. Wang, Z. Sun, M. D. Wang, and T. Xie, "In vivo use of hyperspectral imaging to develop a noncontact endoscopic diagnosis support system for malignant colorectal tumors," Journal of biomedical optics **21**, 16001 (2016).

131. E. S. Hwang, D. Y. Lichtensztajn, S. L. Gomez, B. Fowble, and C. A. Clarke, "Survival after lumpectomy and mastectomy for early stage invasive breast cancer," Cancer **119**, 1402-1411 (2013).

132. L. Jacobs, "Positive Margins: The Challenge Continues for Breast Surgeons," Annals of Surgical Oncology **15**, 1271-1272 (2008).

133. S. V. Panasyuk, S. Yang, D. V. Faller, D. Ngo, R. A. Lew, J. E. Freeman, and A. E. Rogers, "Medical hyperspectral imaging to facilitate residual tumor identification during surgery," Cancer Biology & Therapy **6**, 439-446 (2007).

134. D. R. McCormack, A. J. Walsh, W. Sit, C. L. Arteaga, J. Chen, R. S. Cook, and M. C. Skala, "In vivo hyperspectral imaging of microvessel response to trastuzumab treatment in breast cancer xenografts," Biomedical Optics Express **5**, 2247 (2014).

135. B. Kim, N. Kehtarnavaz, P. LeBoulluec, H. Liu, Y. Peng, and D. Euhus, "Automation of ROI extraction in hyperspectral breast images," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, 2013).

136. R. Pourreza-Shahri, F. Saki, N. Kehtarnavaz, P. Leboulluec, and H. Liu, "Classification of ex-vivo breast cancer positive margins measured by

hyperspectral imaging," in *2013 IEEE International Conference on Image Processing*, (IEEE, 2013).

137. E. Kho, L. L. de Boer, K. K. Van de Vijver, F. van Duijnhoven, M. Vrancken Peeters, H. Sterenborg, and T. J. M. Ruers, "Hyperspectral Imaging for Resection Margin Assessment during Cancer Surgery," Clin Cancer Res **25**, 3572-3580 (2019).

138. E. Kho, B. Dashtbozorg, L. L. de Boer, K. K. Van de Vijver, H. Sterenborg, and T. J. M. Ruers, "Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information," Biomed Opt Express **10**, 4496-4515 (2019).

139. N. Vigneswaran and M. D. Williams, "Epidemiologic Trends in Head and Neck Cancer and Aids in Diagnosis," Oral and Maxillofacial Surgery Clinics of North America **26**, 123-141 (2014).

140. J. Ringash, "Survivorship and Quality of Life in Head and Neck Cancer," J Clin Oncol **33**, 3322-3327 (2015).

141. M. Yao, J. B. Epstein, B. J. Modi, K. B. Pytynia, A. J. Mundt, and L. E. Feldman, "Current surgical treatment of squamous cell carcinoma of the head and neck," Oral Oncol **43**, 213-223 (2007).

142. P. Molony, N. Kharytaniuk, S. Boyle, R. S. R. Woods, G. O'Leary, R. Werner, C. Heffron, L. Feeley, and P. Sheahan, "Impact of positive margins on outcomes of oropharyngeal squamous cell carcinoma according to p16 status," Head & Neck **39**, 1680-1688 (2017).

143. C. Hernández-Vila, "Current prognosis and quality of life following surgical treatment for head and neck squamous cell carcinoma," Plastic and Aesthetic Research **3**, 203 (2016).

144. G. Lu, L. Halig, D. Wang, Z. G. Chen, and B. Fei, "Hyperspectral Imaging for Cancer Surgical Margin Delineation: Registration of Hyperspectral and Histological Images," Proc SPIE Int Soc Opt Eng **9036**, 90360S (2014).

145. G. Lu, D. Wang, X. Qin, L. Halig, S. Muller, H. Zhang, A. Chen, B. W. Pogue, Z. G. Chen, and B. Fei, "Framework for hyperspectral image processing and quantification for cancer detection during animal tumor surgery," Journal of biomedical optics **20**, 126012 (2015).

146. R. Pike, G. Lu, D. Wang, Z. G. Chen, and B. Fei, "A Minimum Spanning Forest-Based Method for Noninvasive Cancer Detection With Hyperspectral Imaging," IEEE Transactions on Biomedical Engineering **63**, 653-663 (2016).

147. B. Fei, G. Lu, X. Wang, H. Zhang, J. V. Little, M. R. Patel, C. C. Griffith, M. W. El-Diery, and A. Y. Chen, "Label-free reflectance hyperspectral imaging for

tumor margin assessment: a pilot study on surgical specimens of cancer patients," Journal of biomedical optics **22**, 7 (2017).

148. Stojan Trajanovski, Caifeng Shan, Pim J. C. Weijtmans, Susan G. Brouwer de Koning, and T. J. M. Ruers, "Tumor Semantic Segmentation in Hyperspectral Images using Deep Learning," in *2nd International Conference on Medical Imaging with Deep Learning*, (2nd International Conference on Medical Imaging with Deep Learning. London, England, UK, 2019).

149. S. G. Brouwer de Koning, P. Weijtmans, M. B. Karakullukcu, C. Shan, E. J. M. Baltussen, L. A. Smit, R. L. P. van Veen, B. H. W. Hendriks, H. Sterenborg, and T. J. M. Ruers, "Toward assessment of resection margins using hyperspectral diffuse reflection imaging (400-1,700 nm) during tongue cancer surgery," Lasers Surg Med (2019).

150. J. L. Jayanthi, G. U. Nisha, S. Manju, E. K. Philip, P. Jeemon, K. V. Baiju, V. T. Beena, and N. Subhash, "Diffuse reflectance spectroscopy: diagnostic accuracy of a non-invasive screening technique for early detection of malignant changes in the oral cavity," BMJ Open **1**, e000071-e000071 (2011).

151. B. Regeling, W. Laffers, A. O. H. Gerstner, S. Westermann, N. A. Müller, K. Schmidt, J. Bendix, and B. Thies, "Development of an image pre-processor for operational hyperspectral laryngeal cancer detection," Journal of Biophotonics **9**, 235-245 (2015).

152. W. Laffers, S. Westermann, B. Regeling, R. Martin, B. Thies, A. O. H. Gerstner, F. Bootz, and N. A. Müller, "Early recognition of cancerous lesions in the mouth and oropharynx: Automated evaluation of hyperspectral image stacks," HNO **64**, 27-33 (2015).

153. Z. Liu, H. Wang, and Q. Li, "Tongue Tumor Detection in Medical Hyperspectral Images," Sensors **12**, 162-174 (2011).

154. D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," Acta Neuropathologica **131**, 803-820 (2016).

155. D. K. Robson, "Pathology & Genetics. Tumours of the Nervous System. World Health Organisation Classification of Tumours. P. Kleihues and k. Cavenee (eds). IARC Press, Lyon, 2000. No. of pages: 314. ISBN: 92 832 2409 4," The Journal of Pathology **193**, 276-276 (2001).

156. N. Sanai and M. S. Berger, "Operative techniques for gliomas and the value of extent of resection," Neurotherapeutics **6**, 478-486 (2009).

157. K. Petrecca, M.-C. Guiot, V. Panet-Raymond, and L. Souhami, "Failure pattern following complete resection plus radiotherapy and temozolomide is at the

resection margin in patients with glioblastoma," Journal of Neuro-Oncology **111**, 19-23 (2012).

158. W. Stummer, J.-C. Tonn, H. M. Mehdorn, U. Nestler, K. Franz, C. Goetz, A. Bink, and U. Pichlmeier, "Counterbalancing risks and gains from extended resections in malignant glioma surgery: a supplemental analysis from the randomized 5-aminolevulinic acid glioma resection study," Journal of Neurosurgery **114**, 613-623 (2011).

159. I. J. Gerard, M. Kersten-Oertel, K. Petrecca, D. Sirhan, J. A. Hall, and D. L. Collins, "Brain shift in neuronavigation of brain tumors: A review," Medical Image Analysis **35**, 403-420 (2017).

160. T. Chowdhury, F. A. Zeiler, G. P. Singh, A. Hailu, H. Loewen, B. Schaller, R. B. Cappellani, and M. West, "The Role of Intraoperative MRI in Awake Neurosurgical Procedures: A Systematic Review," Frontiers in Oncology **8**(2018).

161. A. G. Chacko, N. K. S. Kumar, G. Chacko, R. Athyal, and V. Rajshekhar, "Intraoperative ultrasound in determining the extent of resection of parenchymal brain tumours ? a comparative study with computed tomography and histopathology," Acta Neurochirurgica **145**, 743-748 (2003).

162. M. A. Hammoud, B. L. Ligon, R. Elsouki, W. M. Shi, D. F. Schomer, and R. Sawaya, "Use of intraoperative ultrasound for localizing tumors and determining the extent of resection: a comparative study with magnetic resonance imaging," Journal of Neurosurgery **84**, 737-741 (1996).

163. F. Prada, A. Perin, A. Martegani, L. Aiani, L. Solbiati, M. Lamperti, C. Casali, F. Legnani, L. Mattei, A. Saladino, M. Saini, and F. DiMeco, "Intraoperative Contrast-Enhanced Ultrasound for Brain Tumor Surgery," Neurosurgery **74**, 542-552 (2014).

164. F. W. Floeth, M. Sabel, C. Ewelt, W. Stummer, J. Felsberg, G. Reifenberger, H. J. Steiger, G. Stoffels, H. H. Coenen, and K.-J. Langen, "Comparison of 18F-FET PET and 5-ALA fluorescence in cerebral gliomas," European Journal of Nuclear Medicine and Molecular Imaging **38**, 731-741 (2011).

165. W. Stummer, U. Pichlmeier, T. Meinel, O. D. Wiestler, F. Zanella, and H.-J. Reulen, "Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial," The Lancet Oncology **7**, 392-401 (2006).

166. H. Fabelo, M. Halicek, S. Ortega, M. Shahedi, A. Szolna, J. Piñeiro, C. Sosa, A. O'Shanahan, S. Bisshopp, C. Espino, M. Márquez, M. Hernández, D. Carrera, J. Morera, G. Callico, R. Sarmiento, and B. Fei, "Deep Learning-Based Framework for In Vivo Identification of Glioblastoma Tumor using Hyperspectral Images of Human Brain," Sensors **19**, 920 (2019).

167. H. Fabelo, S. Ortega, R. Lazcano, D. Madronal, M. C. G, E. Juarez, R. Salvador, D. Bulters, H. Bulstrode, A. Szolna, J. F. Pineiro, C. Sosa, J. O. S. A, S. Bisshopp, M. Hernandez, J. Morera, D. Ravi, B. R. Kiran, A. Vega, A. Baez-Quevedo, G. Z. Yang, B. Stanciulescu, and R. Sarmiento, "An Intraoperative Visualization System Using Hyperspectral Imaging to Aid in Brain Tumor Delineation," Sensors (Basel) **18**, E430 (2018).

168. H. Fabelo, S. Ortega, A. Szolna, D. Bulters, J. F. Pineiro, S. Kabwama, A. J-O'Shanahan, H. Bulstrode, S. Bisshopp, B. R. Kiran, D. Ravi, R. Lazcano, D. Madronal, C. Sosa, C. Espino, M. Marquez, M. De La Luz Plaza, R. Camacho, D. Carrera, M. Hernandez, G. M. Callico, J. Morera Molina, B. Stanciulescu, G.-Z. Yang, R. Salvador Perea, E. Juarez, C. Sanz, and R. Sarmiento, "In-Vivo Hyperspectral Human Brain Image Database for Brain Cancer Detection," IEEE Access **7**, 39098-39116 (2019).

169. H. Fabelo, S. Ortega, D. Ravi, B. R. Kiran, C. Sosa, D. Bulters, G. M. Callico, H. Bulstrode, A. Szolna, J. F. Pineiro, S. Kabwama, D. Madronal, R. Lazcano, J. O. S. A, S. Bisshopp, M. Hernandez, A. Baez, G. Z. Yang, B. Stanciulescu, R. Salvador, E. Juarez, and R. Sarmiento, "Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations," PLoS One **13**, e0193721 (2018).

170. B. Fei, G. Lu, M. T. Halicek, X. Wang, H. Zhang, J. V. Little, K. R. Magliocca, M. Patel, C. C. Griffith, M. W. El-Deiry, and A. Y. Chen, "Label-free hyperspectral imaging and quantification methods for surgical margin assessment of tissue specimens of cancer patients," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 4041-4045 (2017).

171. M. Halicek, J. V. Little, X. Wang, C. C. Griffith, M. Patel, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks," Proc SPIE Int Soc Opt Eng **Optical Imaging, Therapeutics, and Advanced Technology in Head and Neck Surgery and Otolaryngology**(2018).

172. M. Halicek, J. V. Little, X. Wang, M. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Tumor margin classification of head and neck cancer using hyperspectral imaging and convolutional neural networks " Proc SPIE Int Soc Opt Eng **10576**, 10576-10574 (2018).

173. S. Ortega, H. Fabelo, D. K. Iakovidis, A. Koulaouzidis, and G. M. Callico, "Use of Hyperspectral/Multispectral Imaging in Gastroenterology. Shedding Some Different Light into the Dark," J Clin Med **8**, E36 (2019).

174. M. Salmivuori, N. Neittaanmaki, I. Polonen, L. Jeskanen, E. Snellman, and M. Gronroos, "Hyperspectral imaging system in the delineation of Ill-defined basal cell carcinomas: a pilot study," J Eur Acad Dermatol Venereol **33**, 71 (2018).

175. C. S. Farah, A. J. Dalley, P. Nguyen, M. Batstone, F. Kordbacheh, J. Perry-Keene, and D. Fielding, "Improved surgical margin definition by narrow band imaging for resection of oral squamous cell carcinoma: A prospective gene expression profiling study," Head Neck **38**, 832-839 (2016).

176. M. Leigh-Ann and R. M. Paul, "MicroRNA: Biogenesis, Function and Role in Cancer," Current Genomics **11**, 537-561 (2010).

177. C. S. Farah, "Narrow Band Imaging-guided resection of oral cavity cancer decreases local recurrence and increases survival," Oral Dis **24**, 89-97 (2018).

178. C. S. Farah, S. A. Fox, and A. J. Dalley, "Integrated miRNA-mRNA spatial signature for oral squamous cell carcinoma: a prospective profiling study of Narrow Band Imaging guided resection," Sci Rep **8**, 823 (2018).

179. M. Halicek, J. V. Little, X. Wang, A. Y. Chen, and B. Fei, "Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks," Journal of biomedical optics **24**, 1-9 (2019).

180. F. Manni, F. van der Sommen, S. Zinger, E. Kho, S. G. Brouwer de Koning, T. J. M. Ruers, C. Shan, J. Schleipen, P. H. N. de With, B. Fei, and C. A. Linte, "Automated tumor assessment of squamous cell carcinoma on tongue cancer patients with hyperspectral imaging," Proc SPIE 10951 **Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling**, 109512K (2019).

181. P. J. C. Weijtmans, C. Shan, T. Tan, S. G. Brouwer de Koning, and T. J. M. Ruers, "A Dual Stream Network for Tumor Detection in Hyperspectral Images," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI) (2019).

182. Martin Halicek, Himar Fabelo, Samuel Ortega, James V. Little, Xu Wang, Amy Y. Chen, Gustavo Marrero Callico, Larry Myers, Baran D. Sumer, and B. Fei, "Cancer margin evaluation using machine learning in hyperspectral images of head and neck cancer," Proc SPIE Int Soc Opt Eng **Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling**(2019).

183. G. Lu, X. Qin, D. Wang, S. Muller, H. Zhang, A. Chen, Z. G. Chen, and B. Fei, "Hyperspectral Imaging of Neoplastic Progression in a Mouse Model of Oral Carcinogenesis," Proc SPIE Int Soc Opt Eng **Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging**(2016).

184. M. Halicek, J. V. Little, X. Wang, Z. G. Chen, M. Patel, C. C. Griffith, M. W. El-Deiry, N. F. Saba, A. Y. Chen, and B. Fei, "Deformable registration of histological cancer margins to gross hyperspectral images using demons," Proc SPIE Int Soc Opt Eng **10581**, 1-8 (2018).

185. L. G. H. Derek, G. B. Philipp, H. Mark, and J. H. David, "Medical image registration," Physics in Medicine & Biology **46**, R1 (2001).

186. W. R. Crum, T. Hartkens, and D. L. Hill, "Non-rigid image registration: theory and practice," The British journal of radiology **77 Spec No 2**, S140-153 (2004).

187. J. P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," Medical Image Analysis **2**, 243-260 (1998).

188. B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artif. Intell. **17**, 185-203 (1981).

189. L. A. Sordillo, Y. Pu, S. Pratavieira, Y. Budansky, and R. R. Alfano, "Deep optical imaging of tissue using the second and third near-infrared spectral windows," Journal of biomedical optics **19**, 056004 (2014).

190. C. Ash, M. Dubec, K. Donne, and T. Bashford, "Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods," Lasers Med Sci **32**, 1909-1918 (2017).

191. H. Zhang, D. Salo, D. M. Kim, S. Komarov, Y. C. Tai, and M. Y. Berezin, "Penetration depth of photons in biological tissues from hyperspectral imaging in shortwave infrared in transmission and reflection geometries," Journal of biomedical optics **21**, 126006 (2016).

192. Martín Abadi, Ashish Agarwal, Paul Barham, and e. a. Eugene Brevdo, "TensorFlow: Large-scale machine learning on heterogeneous systems," https://www.tensorflow.org (2015).

193. G. Florimbi, H. Fabelo, E. Torti, R. Lazcano, D. Madronal, S. Ortega, R. Salvador, F. Leporati, G. Danese, A. Baez-Quevedo, G. M. Callico, E. Juarez, C. Sanz, and R. Sarmiento, "Accelerating the K-Nearest Neighbors Filtering Algorithm to Optimize the Real-Time Classification of Human Brain Tumor in Hyperspectral Images," Sensors (Basel) **18**, E2314 (2018).

194. R. Lazcano, D. Madroñal, R. Salvador, K. Desnos, M. Pelcat, R. Guerra, H. Fabelo, S. Ortega, S. Lopez, G. M. Callico, E. Juarez, and C. Sanz, "Porting a PCA-based hyperspectral image dimensionality reduction algorithm for brain cancer detection on a manycore architecture," Journal of Systems Architecture **77**, 101-111 (2017).

195. D. Madroñal, R. Lazcano, R. Salvador, H. Fabelo, S. Ortega, G. M. Callico, E. Juarez, and C. Sanz, "SVM-based real-time hyperspectral image classifier on a manycore architecture," Journal of Systems Architecture **80**, 30-40 (2017).

196. H. M. Baddour, Jr., K. R. Magliocca, and A. Y. Chen, "The importance of margins in head and neck cancer," J Surg Oncol **113**, 248-255 (2016).

197. D. Zanoni, J. C. Migliacci, B. Xu, and et al., "A proposal to redefine close surgical margins in squamous cell carcinoma of the oral tongue," JAMA Otolaryngol Head Neck Surg **143**, 555-560 (2017).

198. A. Liu and X. Xu, "MicroRNA Isolation from Formalin-Fixed, Paraffin-Embedded Tissues," in *Formalin-Fixed Paraffin-Embedded Tissues: Methods and Protocols*, F. Al-Mulla, ed. (Humana Press, Totowa, NJ, 2011), pp. 259-267.

199. S. Marur and A. A. Forastiere, "Head and Neck Squamous Cell Carcinoma: Update on Epidemiology, Diagnosis, and Treatment," Mayo Clinic Proceedings **91**, 386-396 (2016).

200. A. Bozec, D. Culié, G. Poissonnet, and O. Dassonville, "Current role of primary surgical treatment in patients with head and neck squamous cell carcinoma," Current Opinion in Oncology **31**, 138-145 (2019).

201. S. van Keulen, N. Nishio, S. Fakurnejad, A. Birkeland, B. A. Martin, G. Lu, Q. Zhou, S. U. Chirita, T. Forouzanfar, A. D. Colevas, N. S. van den Berg, and E. L. Rosenthal, "The Clinical Application of Fluorescence-Guided Surgery in Head and Neck Cancer," J Nucl Med **60**, 758-763 (2019).

202. A. O. H. Gerstner, "Early detection in head and neck cancer – current state and future perspectives," GMS Curr Top Otorhinolaryngol Head Neck Surg **7**(2010).

203. L. J. DiNardo, J. Lin, L. S. Karageorge, and C. N. Powers, "Accuracy, Utility, and Cost of Frozen Section Margins in Head and Neck Cancer Surgery," The Laryngoscope **101**, 1773-1776 (2000).

204. N. F. Ribeiro, D. R. Godden, G. E. Wilson, D. M. Butterworth, and R. T. Woodwards, "Do frozen sections help achieve adequate surgical margins in the resection of oral carcinoma?," Int J Oral Maxillofac Surg **32**, 152-158 (2003).

205. E. Du, T. J. Ow, Y. T. Lo, A. Gersten, B. A. Schiff, A. B. Tassler, and R. V. Smith, "Refining the utility and role of Frozen section in head and neck squamous cell carcinoma resection," Laryngoscope **126**, 1768-1775 (2016).

206. E. M. Layfield, R. L. Schmidt, M. Esebua, and L. J. Layfield, "Frozen Section Evaluation of Margin Status in Primary Squamous Cell Carcinomas of the Head and Neck: A Correlation Study of Frozen Section and Final Diagnoses," Head Neck Pathol **12**, 175-180 (2018).

207. C. Black, J. Marotti, E. Zarovnaya, and J. Paydarfar, "Critical evaluation of frozen section margins in head and neck cancer resections," Cancer **107**, 2792-2800 (2006).

208. R. W. Gao, N. T. Teraphongphom, N. S. van den Berg, B. A. Martin, N. J. Oberhelman, V. Divi, M. J. Kaplan, S. S. Hong, G. Lu, R. Ertsey, W. Tummers, A. J. Gomez, F. C. Holsinger, C. S. Kong, A. D. Colevas, J. M. Warram, and E. L. Rosenthal, "Determination of Tumor Margins with Surgical Specimen Mapping Using Near-Infrared Fluorescence," Cancer Res **78**, 5144-5154 (2018).

209. M. G. Crowson, J. Ranisau, A. Eskander, A. Babier, B. Xu, R. R. Kahmke, J. M. Chen, and T. C. Y. Chan, "A contemporary review of machine learning in otolaryngology-head and neck surgery," Laryngoscope **00**, 1-7 (2019).

210. B. Fei, G. Lu, X. Wang, H. Zhang, J. V. Little, M. R. Patel, C. C. Griffith, M. W. El-Diery, and A. Y. Chen, "Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients," Journal of biomedical optics **22**, 1 (2017).

211. M. Halicek, H. Fabelo, S. Ortega, G. M. Callico, and B. Fei, "In-Vivo and Ex-Vivo Tissue Analysis through Hyperspectral Imaging Techniques: Revealing the Invisible Features of Cancer," Cancers (Basel) **11**(2019).

212. J. Shapey, Y. Xie, E. Nabavi, R. Bradford, S. R. Saeed, S. Ourselin, and T. Vercauteren, "Intraoperative multispectral and hyperspectral label-free imaging: A systematic review of in vivo clinical studies," J Biophotonics, e201800455 (2019).

213. A. Hellebust, K. Rosbach, J. K. Wu, J. Nguyen, A. Gillenwater, N. Vigneswaran, and R. Richards-Kortum, "Vital-dye-enhanced multimodal imaging of neoplastic progression in a mouse model of oral carcinogenesis," Journal of biomedical optics **18**, 126017 (2013).

214. N. Thekkek, T. Muldoon, A. D. Polydorides, D. M. Maru, N. Harpaz, M. T. Harris, W. Hofstettor, S. P. Hiotis, S. A. Kim, A. J. Ky, S. Anandasabapathy, and R. Richards-Kortum, "Vital-dye enhanced fluorescence imaging of GI mucosa: metaplasia, neoplasia, inflammation," Gastrointestinal Endoscopy **75**, 877-887 (2012).

215. M. Halicek, M. Shahedi, J. V. Little, A. Y. Chen, L. L. Myers, B. D. Sumer, and B. Fei, *Detection of squamous cell carcinoma in digitized histological images from the head and neck using convolutional neural networks*, Proceedings Volume 10956, Medical Imaging 2019: Digital Pathology (SPIE, San Diego, CA, USA, 2019), Vol. 10956, p. 109560K

216. M. Halicek, H. Fabelo, S. Ortega, J. V. Little, X. Wang, A. Y. Chen, G. M. Callico, L. Myers, B. D. Sumer, and B. Fei, "Hyperspectral imaging for head and neck cancer detection: specular glare and variance of the tumor margin in surgical specimens," Journal of Medical Imaging **6**, 035004 (2019).

217. S. I. Christian Szegedy, Vincent Vanhoucke, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," arXiv: Computational Research Repository **1602.07261**(2016).

218. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," Proc IEEE Int Conf Comput Vis, 618-626 (2017).

219. Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced CNN-Based Hyperspectral Recovery from RGB Images," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) **2473-9944**, 1052-1061 (2018).

220. C. La Vecchia, M. Malvezzi, C. Bosetti, W. Garavello, P. Bertuccio, F. Levi, and E. Negri, "Thyroid cancer mortality and incidence: a global overview," Int J Cancer **136**, 2187-2195 (2015).

221. "Thyroid Cancer" (American Cancer Society, 2019), retrieved October 7, 2019, 2019, cancer.org/cancer/thyroid-cancer.

222. R. V. Lloyd, D. Buehler, and E. Khanafshar, "Papillary Thyroid Carcinoma Variants," Head and Neck Pathology **5**, 51-56 (2011).

223. M. Sobrinho-Simoes, C. Eloy, J. Magalhaes, C. Lobo, and T. Amaro, "Follicular thyroid carcinoma," Mod Pathol **24 Suppl 2**, S10-18 (2011).

224. M. Roy, H. Chen, and R. S. Sippel, "Current Understanding and Management of Medullary Thyroid Cancer," The Oncologist **18**, 1093-1100 (2013).

225. B. R. Haugen, E. K. Alexander, K. C. Bible, G. M. Doherty, S. J. Mandel, Y. E. Nikiforov, F. Pacini, G. W. Randolph, A. M. Sawka, M. Schlumberger, K. G. Schuff, S. I. Sherman, J. A. Sosa, D. L. Steward, R. M. Tuttle, and L. Wartofsky, "2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer," Thyroid **26**, 1-133 (2015).

226. R. Mallick, T. M. Stevens, T. S. Winokur, A. Asban, T. N. Wang, B. M. Lindeman, J. R. Porterfield, and H. Chen, "Is Frozen-Section Analysis During Thyroid Operation Useful in the Era of Molecular Testing?," Journal of the American College of Surgeons **228**, 474-479 (2019).

227. R. W. Berg, T. W. Yen, D. B. Evans, B. Hunt, F. A. Quiroz, S. D. Wilson, and T. S. Wang, "Analysis of an institutional protocol for thyroid lobectomy: Utility of routine intraoperative frozen section and expedited (overnight) pathology," Surgery **159**, 512-517 (2016).

228. H. Boukheris, R. E. Curtis, C. E. Land, and G. M. Dores, "Incidence of Carcinoma of the Major Salivary Glands According to the WHO Classification, 1992 to 2006: A Population-Based Study in the United States," Cancer Epidemiology Biomarkers & Prevention **18**, 2899-2906 (2009).

229. V. S. H. To, J. Y. W. Chan, R. K. Y. Tsang, and W. I. Wei, "Review of Salivary Gland Neoplasms," ISRN Otolaryngology **2012**, 6 (2012).

230. S. Bobati, B. Patil, and V. Dombale, "Histopathological study of salivary gland tumors," Journal of Oral and Maxillofacial Pathology **21**, 46-50 (2017).

231. K. Chatura, "Polymorphous low grade adenocarcinoma," Journal of Oral and Maxillofacial Pathology **19**, 77-82 (2015).

232. M. Suzuki, Y. Nakaegawa, T. Kobayashi, T. Kawase, T. Matsuzuka, and S. Murono, "The role of frozen section biopsy for parotid gland tumour with benign fine-needle aspiration cytology," The Journal of Laryngology & Otology **133**, 227-229 (2019).

233. D. C. Howlett, E. Skelton, and A. B. Moody, "Establishing an accurate diagnosis of a parotid lump: evaluation of the current biopsy methods – fine needle aspiration cytology, ultrasound-guided core biopsy, and intraoperative frozen section," British Journal of Oral and Maxillofacial Surgery **53**, 580-583 (2015).

234. S. Grasl, L. Kadletz, S. Janik, A. Riedl, B. Erlacher, M. Formanek, M. C. Grasl, and B. M. Erovic, "Fine-needle aspiration cytology and intraoperative frozen section in parotid gland tumour surgery: A retrospective multicenter analysis of 417 cases," Clinical Otolaryngology **44**, 461-465 (2019).

235. B. Jansen-Winkeln, N. Holfert, H. Kohler, Y. Moulla, J. P. Takoh, S. M. Rabe, M. Mehdorn, M. Barberio, C. Chalopin, T. Neumuth, and I. Gockel, "Determination of the transection margin during colorectal resection with hyperspectral imaging (HSI)," Int J Colorectal Dis **34**, 731-739 (2019).

236. B. Fei, G. Lu, X. Wang, H. Zhang, J. V. Little, M. R. Patel, C. C. Griffith, M. W. El-Diery, and A. Y. Chen, "Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients," Journal of biomedical optics **22**, 1-7 (2017).

237. M. Barberio, M. Maktabi, I. Gockel, N. Rayes, B. Jansen-Winkeln, H. Köhler, S. M. Rabe, L. Seidemann, J. P. Takoh, M. Diana, T. Neumuth, and C. Chalopin, "Hyperspectral based discrimination of thyroid and parathyroid during surgery," Current Directions in Biomedical Engineering **4**, 399-402 (2018).

238. E. L. Wisotzky, F. C. Uecker, P. Arens, S. Dommerich, A. Hilsmann, and P. Eisert, "Intraoperative hyperspectral determination of human tissue properties," Journal of biomedical optics **23**, 1-8, 8 (2018).

239. M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, L. Myers, B. D. Sumer, and B. Fei, "Hyperspectral Imaging of Head and Neck Squamous Cell Carcinoma for Cancer Margin Detection in Surgical Specimens from 102 Patients Using Deep Learning," Cancers **11**, 1367 (2019).

240. J. J. Vos, "Colorimetric and photometric properties of a 2° fundamental observer," Color Research & Application **3**, 125-128 (1978).

241. E. Berra, S. Gibson-Poole, A. MacArthur, R. Gaulton, and A. Hamilton, "Eestimation of the spectral sensitivity functions of un-modified and modified commercial off-the-shelf digital cameras to enable their use as a multispectral imaging system for UAVs," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **XL-1/W4**, 207-214 (2015).

242. C. Mauer and D. Wueller, *Measuring the spectral response with a set of interference filters*, IS&T/SPIE Electronic Imaging (SPIE, 2009), Vol. 7250.

243. M. Halicek, M. Shahedi, J. V. Little, A. Y. Chen, L. L. Myers, B. D. Sumer, and B. Fei, "Head and Neck Cancer Detection in Digitized Whole-Slide Histology Using Convolutional Neural Networks," Scientific Reports **9**, 14043 (2019).

244. J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," arXiv: Computational Research Repository **1709.01507**(2018).

245. V. Ajila, H. Shetty, S. Babu, V. Shetty, and S. Hegde, "Human Papilloma Virus Associated Squamous Cell Carcinoma of the Head and Neck," J Sex Transm Dis **2015**, 791024 (2015).

246. M. G. Muller, T. A. Valdez, I. Georgakoudi, V. Backman, C. Fuentes, S. Kabani, N. Laver, Z. Wang, C. W. Boone, R. R. Dasari, S. M. Shapshay, and M. S. Feld, "Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma," Cancer **97**, 1681-1692 (2003).

247. D. Roblyer, R. Richards-Kortum, K. Sokolov, A. K. El-Naggar, M. D. Williams, C. Kurachi, and A. M. Gillenwater, "Multispectral optical imaging device for in vivo detection of oral neoplasia," Journal of biomedical optics **13**, 024019 (2008).

248. B. Swinson, W. Jerjes, M. El-Maaytah, P. Norris, and C. Hopper, "Optical techniques in diagnosis of head and neck malignancy," Oral Oncol **42**, 221-228 (2006).

249. H. W. Beumer, K. Vishwanath, L. Puscas, H. R. Afshari, N. Ramanujam, and W. T. Lee, "Detection of squamous cell carcinoma and corresponding biomarkers using optical spectroscopy," Otolaryngol Head Neck Surg **144**, 390-394 (2011).

250. A. M. Schreiner and G. C. H. Yang, "Adenomatoid nodules are the main cause for discrepant histology in 234 thyroid fine-needle aspirates reported as follicular neoplasm," Diagnostic Cytopathology **40**, 375-379 (2012).

251. N. Dabelic, N. Matesa, D. Matesa-Anic, and Z. Kusic, "Malignancy risk assessment in adenomatoid nodules and suspicious follicular lesions of the thyroid obtained by fine needle aspiration cytology," Collegium antropologicum **34**, 349-354 (2010).

252. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International*

Conference on Artificial Intelligence and Statistics, T. Yee Whye and T. Mike, eds. (PMLR, Proceedings of Machine Learning Research, 2010), pp. 249--256.

253. M. D. Zeiler, "ADADELTA: An adaptive learning rate method," arXiv: Computational Research Repository **121.5701**(2012).

254. E. Mahe, S. Ara, M. Bishara, A. Kurian, S. Tauqir, N. Ursani, P. Vasudev, T. Aziz, C. Ross, and A. Lytwyn, "Intraoperative pathology consultation: error, cause and impact," Can J Surg **56**, E13-18 (2013).

255. G. Dionigi, P. Castano, V. Bertolini, L. Boni, F. Rovera, M. L. Tanda, C. Capella, L. Bartalena, and R. Dionigi, "Simultaneous medullary and papillary thyroid cancer: two case reports," Journal of Medical Case Reports **1**, 133 (2007).

256. Z. Adnan, E. Arad, J. Dana, Y. Shendler, and E. Baron, "Simultaneous occurrence of medullary and papillary thyroid microcarcinomas: a case series and review of the literature," Journal of Medical Case Reports **7**, 26 (2013).

257. H. Mazeh, A. Orlev, I. Mizrahi, D. J. Gross, and H. R. Freund, "Concurrent Medullary, Papillary, and Follicular Thyroid Carcinomas and Simultaneous Cushing's Syndrome," European Thyroid Journal **4**, 65-68 (2015).

258. K. K. Tasche, M. R. Buchakjian, N. A. Pagedar, and S. M. Sperry, "Definition of “close margin” in oral cancer surgery and association of margin distance with local recurrence rate," JAMA Otolaryngol Head Neck Surg (2017).

259. X. Fuyong, X. Yuanpu, S. Hai, L. Fujun, and Y. Lin, "Deep Learning in Microscopy Image Analysis: A Survey," IEEE Trans Neural Netw Learn Syst **29**, 4550-4568 (2018).

260. N. Linder, "Identification of tumor epithelium and stroma in tissue microarrays using texture analysis," Diagn Pathol **7**(2012).

261. F. Bianconi, A. Álvarez-Larrán, and A. Fernández, "Discrimination between tumour epithelium and stroma via perception-based features," Neurocomputing **154**, 119-126 (2015).

262. J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images," Neurocomputing **191**, 214-223 (2016).

263. L. a. S. Hou, Dimitris and Kurc, Tahsin M. and Gao, Yi and Davis, James E. and Saltz, Joel H., "Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification," Proc IEEE Conf Comp Vis Pattern Recogn (2016).

264. Z.-H. Zhou, Y. Jiang, Y.-B. Yang, and S.-F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," Artif Intell Med **24**, 25-36 (2002).

265. G. Carneiro, T. Peng, C. Bayer, and N. Navab, "Weakly-Supervised Structured Output Learning with Flexible and Latent Graphs Using High-Order Loss Functions," in *Proc IEEE Int Conf Comput Vis*, 2015), 648-656.

266. E. Kim, Z. Baloch, and C. Kim, "Computer assisted detection and analysis of tall cell variant papillary thyroid carcinoma in histological images," in *Proc SPIE Int Soc Opt Eng*, 2015), 8.

267. l. J. Angel and M. A. Rajam, "Automatic classification of thyroid histopathology images using multi-classifier system," Multimed Tools Appl **76**, 18711-18730 (2017).

268. W. Wang, J. A. Ozolek, and G. K. Rohde, "Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images," Cytometry A **77A**, 485-494 (2010).

269. B. Gopinath and N. Shanthi, "Computer-aided diagnosis system for classifying benign and malignant thyroid nodules in multi-stained FNAB cytological images," Australas Phys Eng Sci Med **36**, 219-230 (2013).

270. A. Daskalakis, S. Kostopoulos, P. Spyridonos, D. Glotsos, P. Ravazoula, M. Kardari, I. Kalatzis, D. Cavouras, and G. Nikiforidis, "Design of a multi-classifier system for discriminating benign from malignant thyroid nodules using routinely H&E-stained cytological images," Comput Biol Med **38**, 196-203 (2008).

271. J. A. Ozolek, A. B. Tosun, W. Wang, C. Chen, S. Kolouri, S. Basu, H. Huang, and G. K. Rohde, "Accurate diagnosis of thyroid follicular lesions from nuclear morphology using supervised learning," Med Image Anal **18**, 772-780 (2014).

272. L. Pantanowitz, J. H. Sinard, W. H. Henricks, L. A. Fatheree, A. B. Carter, L. Contis, B. A. Beckwith, A. J. Evans, A. Lal, and A. V. Parwani, "Validating Whole Slide Imaging for Diagnostic Purposes in Pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center," Archives of pathology & laboratory medicine **137**, 1710-1722 (2013).

273. S. Mukhopadhyay, M. D. Feldman, E. Abels, R. Ashfaq, S. Beltaifa, N. G. Cacciabeve, H. P. Cathro, L. Cheng, K. Cooper, G. E. Dickey, R. M. Gill, R. P. J. Heaton, R. Kerstens, G. M. Lindberg, R. K. Malhotra, J. W. Mandell, E. D. Manlucu, A. M. Mills, S. E. Mills, C. A. Moskaluk, M. Nelis, D. T. Patil, C. G. Przybycin, J. P. Reynolds, B. P. Rubin, M. H. Saboorian, M. Salicru, M. A. Samols, C. D. Sturgis, K. O. Turner, M. R. Wick, J. Y. Yoon, P. Zhao, and C. R. Taylor, "Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study)," Am J Surg Path **42**, 39-52 (2018).

274. K. G. Yun Liu, Mohammad Norouzi, George E. Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q. Nelson, Greg S. Corrado, Jason D. Hipp, Lily Peng, Martin C. Stumpe, "Detecting Cancer

Metastases on Gigapixel Pathology Images," arXiv: Computational Research Repository **abs/1703.02442**(2017).

275. B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, J. van der Laak, C. C. the, M. Hermsen, Q. F. Manson, M. Balkenhol, O. Geessink, N. Stathonikos, M. C. van Dijk, P. Bult, F. Beca, A. H. Beck, D. Wang, A. Khosla, R. Gargeya, H. Irshad, A. Zhong, Q. Dou, Q. Li, H. Chen, H. J. Lin, P. A. Heng, C. Hass, E. Bruni, Q. Wong, U. Halici, M. U. Oner, R. Cetin-Atalay, M. Berseth, V. Khvatkov, A. Vylegzhanin, O. Kraus, M. Shaban, N. Rajpoot, R. Awan, K. Sirinukunwattana, T. Qaiser, Y. W. Tsang, D. Tellez, J. Annuscheit, P. Hufnagl, M. Valkonen, K. Kartasalo, L. Latonen, P. Ruusuvuori, K. Liimatainen, S. Albarqouni, B. Mungal, A. George, S. Demirci, N. Navab, S. Watanabe, S. Seno, Y. Takenaka, H. Matsuda, H. Ahmady Phoulady, V. Kovalev, A. Kalinovsky, V. Liauchuk, G. Bueno, M. M. Fernandez-Carrobles, I. Serrano, O. Deniz, D. Racoceanu, and R. Venancio, "Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer," JAMA **318**, 2199-2210 (2017).

276. A. K. Dayong Wang, Rishab Gargeya, Humayun Irshad, Andrew H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," arXiv: Computational Research Repository **1606.05718**(2016).

277. P. H. N. de With, S. Zinger, and F. Ghazvinian Zanjani, "Cancer detection in histopathology whole-slide images using conditional random fields on deep embedded spaces," Proc SPIE Int Soc Opt Eng **2015810I**, 17 (2018).

278. G. Litjens, P. Bandi, B. Ehteshami Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, Q. F. Manson, N. Stathonikos, A. Baidoshvili, P. van Diest, C. Wauters, M. van Dijk, and J. van der Laak, "1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset," Gigascience **7**(2018).

279. W. Bulten, P. Bandi, J. Hoven, R. V. Loo, J. Lotz, N. Weiss, J. V. Laak, B. V. Ginneken, C. Hulsbergen-van de Kaa, and G. Litjens, "Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard," Sci Rep **9**, 864 (2019).

280. A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. C. Shih, J. Tomaszewski, F. A. Gonzalez, and A. Madabhushi, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent," Sci Rep **7**, 46450 (2017).

281. A. Gertych, Z. Swiderska-Chadaj, Z. Ma, N. Ing, T. Markiewicz, S. Cierniak, H. Salemi, S. Guzman, A. E. Walts, and B. S. Knudsen, "Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides," Sci Rep **9**, 1483 (2019).

282. L. Nguyen, A. B. Tosun, J. L. Fine, A. V. Lee, D. L. Taylor, and S. C. Chennubhotla, "Spatial Statistics for Segmenting Histological Structures in H&E Stained Tissue Images," IEEE Trans Med Imaging **36**, 1522-1532 (2017).

283. D. Tellez, M. Balkenhol, I. Otte-Holler, R. van de Loo, R. Vogels, P. Bult, C. Wauters, W. Vreuls, S. Mol, N. Karssemeijer, G. Litjens, J. van der Laak, and F. Ciompi, "Whole-Slide Mitosis Detection in H&E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks," IEEE Trans Med Imaging (2018).

284. J. W. Wei, L. J. Tafe, Y. A. Linnik, L. J. Vaickus, N. Tomita, and S. Hassanpour, "Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks," Sci Rep **9**, 3358 (2019).

285. S. I. Christian Szegedy, Vincent Vanhoucke, Jonathon Shlens, Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision," arXiv: Computational Research Repository **1512.00567**(2015).

286. T. H. Gasparoto, C. E. de Oliveira, L. T. de Freitas, C. R. Pinheiro, R. N. Ramos, A. L. da Silva, G. P. Garlet, J. S. da Silva, and A. P. Campanelli, "Inflammatory events during murine squamous cell carcinoma development," J Inflamm **9**, 46-46 (2012).

287. T. K. Lee, J. F. Silverman, R. D. Horner, and C. W. Scarantino, "Overlap of nuclear diameters in lung cancer cells," Analytical and quantitative cytology and histology **12**, 275-278 (1990).

288. P. Liu, Z. Zhu, C. Zeng, and G. Nie, "Specific absorption spectra of hemoglobin at different PO2 levels: potential noninvasive method to detect PO2 in tissues," Journal of biomedical optics **12**, 125002 (2012).

289. M. G. Vander Heiden, L. C. Cantley, and C. B. Thompson, "Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation," Science **324**, 1029 (2009).

290. A. E. Cerussi, N. S. Shah, D. Hsiang, A. Durkin, J. A. Butler, and B. J. Tromberg, "In vivo absorption, scattering, and physiologic properties of 58 malignant breast tumors determined by broadband diffuse optical spectroscopy," Journal of biomedical optics **11**, 1-6, 6 (2006).

291. P.-H. C. Chen, K. Gadepalli, R. MacDonald, Y. Liu, S. Kadowaki, K. Nagpal, T. Kohlberger, J. Dean, G. S. Corrado, J. D. Hipp, C. H. Mermel, and M. C. Stumpe, "An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis," Nature Medicine **25**, 1453-1457 (2019).

292. G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. Werneck Krauss Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs,

"Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," Nature Medicine **25**, 1301-1309 (2019).

293. M. Bryne, H. Koppang, R. Lilleng, and A. Kjaerheim, "Malignancy grading of the deep invasive margins of oral squamouc cell caricnomas has high prognostic value," Journal of Pathology **166**, 375-381 (1992).

294. S. Fleskens and P. Slootweg, "Grading systems in head and neck dysplasia: their prognostic value, weaknesses and utility," Head Neck Oncol **1**, 11 (2009).

295. F. A. Sawair, C. R. Irwin, D. J. Gordon, A. G. Leonard, and M. Stephenson, "Invasive front grading: reliability and usefulness in the management of oral squamous cell carcinoma," J Oral Pathol Med **32**, 1-9 (2003).

296. M. Brandwein-Gensler, R. V. Smith, B. Wang, C. Penner, A. Theilken, D. Broughel, B. Schiff, R. P. Owen, J. Smith, C. Sarta, T. Hebert, R. Nason, M. Ramer, M. DeLacure, D. Hirsch, D. Myssiorek, K. Heller, M. Prystowsky, N. F. Schlecht, and A. Negassa, "Validation of the Histologic Risk Model in a New Cohort of Patients With Head and Neck Squamous Cell Carcinoma," The American Journal of Surgical Pathology **34**, 676-688 (2010).

297. T. Banzato, G. B. Cherubini, M. Atzori, and A. Zotti, "Development of a deep convolutional neural network to predict grading of canine meningiomas from magnetic resonance images," Vet J **235**, 90-92 (2018).

298. P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, and E. al., "Predicting cancer outcomes from histology and genomics using convolutional networks," PNAS **115**, E2970–E2979 (2017).

299. W. D. Smith, R. C. Dutton, and N. T. Smith, "A measure of association for assessing prediction accuracy that is a generalization of non-parametric ROC area," Stat Med **15**, 1199-1215 (1996).

300. M. L. McHugh, "Interrater reliability: the kappa statistic," Biochemia Medica **22**, 276-282 (2012).

301. E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. J. Wild, J. H. Rüschoff, and M. Claassen, "Automated Gleason grading of prostate cancer tissue microarrays via deep learning," Scientific reports **8**, 12054-12054 (2018).

302. H. D. Couture, L. A. Williams, J. Geradts, S. J. Nyante, E. N. Butler, J. S. Marron, C. M. Perou, M. A. Troester, and M. Niethammer, "Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype," NPJ Breast Cancer **4**, 30 (2018).